Jan Selén

Interdependent Systems with Serially Correlated Errors

Selected publications, vol. 34

Uppsala Universitet
Statistiska Institutionen

University of Uppsala
Department of Statistics
P.O. Box 2300
S–750 02 Uppsala 2 Sweden

1975

University of Uppsala
Department of Statistics
Selected publications, vol. 34

Jan Selén

Interdependent Systems with Serially Correlated Errors

UNIV.
LEIDEN
BIBL.

©Jan Selén

ISBN 91–506–0015–X

Typed by
TEXTGRUPPEN I UPPSALA AB

Printed by
LIBER-TRYCK, STOCKHOLM, 1975

CONTENTS

Abstract		V
Preface		VI
Principles of notation		VIII
Chapter 1:	**A General Introduction to the Subject**	9
1.1	Simultaneity and serial correlation	9
1.2	The consequences of serial correlation	10
1.3	The presence of autocorrelated disturbances in econometric models	11
1.4	The specification of a disturbance pattern	15
1.5	Some suggested estimators	16
Chapter 2:	**The specification of the model**	24
2.1	The basic system	24
2.2	The transformed system	30
2.3	The identification problem: Some possible approaches	33
2.4	Testing for serial correlation	36
Chapter 3:	**The full-information maximum likelihood estimator**	44
3.1	Introduction	44
3.2	The FIMLA estimator	45
3.3	Restrictions and the asymptotic distribution	48
3.4	Some further points	53
Chapter 4:	**Structural and reduced form oriented estimators**	55
4.1	Introduction	55
4.2	Two limited-information methods	56
4.3	Three full-information methods	59
4.4	The consequences of an estimated disturbance covariance matrix	70
4.5	The gain in efficiency	73
Chapter 5:	**An extension of structural and reduced form oriented estimators**	77 77
5.1	Introduction and two principles	77
5.2	The calculated residuals as regressors approach	79
5.3	The GLS-approach I: Limited-Information methods	84
5.4	The GLS-approach II: Full-Information methods	88

Chapter 6: Computational methods 96
6.1 Introduction 96
6.2 The Davidon-Fletcher-Powell method 96
6.3 Fix-point- and related methods 97
6.4 Multiple solutions 103

Chapter 7: Some Monte-Carlo experiments 104
7.1 Introduction 104
7.2 Experimental design 105
7.3 Results 108

Chapter 8: Estimation of Klein's model I 130
8.1 The model 130
8.2 Results 131

References 140

ABSTRACT

Selén Jan: Interdependent Systems with Serially Correlated Errors. The main subject dealt with in this monograph, is estimation of interdependent systems with serially correlated disturbances. In a general introduction the frequency, the sources and the consequences of serially correlated errors are discussed. Some approaches for specifying a process generating the disturbances and suggested modifications of different estimation methods are given. Different approaches to the identification problem are recapitulated and tests for serial correlation examined.

The base for the development of estimation methods, is the assumption that the disturbances are generated by a first order vector autoregressive process. The full-information maximum likelihood estimator is the first method considered. This method, with the associated asymptotic covariance matrix, is derived for a general case.

Two new classes of estimators are suggested in the subsequent section. The reduced form oriented and the structural form oriented methods are derived utilizing a respecification of an interdependent system. Two and three stage least squares are obtained as members of the first class; the GEID estimator belongs to the second. The structural form oriented methods can be regarded as suggestions for the problem of undersized samples. Both classes are extended to the case of serially correlated errors. The asymptotic distributions of full- and limited-information members are given. Computational methods are presented before the results of Monte Carlo experiments are discussed. It is shown that the gain in using the methods accounting for serial correlation can be considerable. Finally, the methods developed are applied to Klein's model I.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548914_0017

PREFACE

Many individuals have contributed to what is good in this study. The focus on estimation aspects of interdependent systems was inspired by the large research project "Icke-lineära kunskapsmodeller. Nipals estimation jämte tillämpningar " initiated by Herman Wold. Ejnar Lyttkens has been my teacher. His interest and knowledge has been of great importance to my work. Anders Christoffersson was my able teacher during Lyttkens' visit to the United States. Many helpful and clarifying discussions have been carried on with Anders Ågren. I wish to express my gratitude to these three persons.

My thanks are extended to Thomas Berglund, David Edgerton and other present and former members of the Institute of Statistics and the Econometric group, Uppsala, for rewarding discussions. Valuable suggestions on the English text were made by Kerstin Lönnqvist. Ingrid Björklund, Karin Ingemarsson and Ann-Christine Pettersson had the patience typing my manuscript.

Uppsala, March 1975 Jan Selén

PRINCIPLES OF NOTATION

A main principle of notation is that Greek letters are used for unobservable quantities while Roman letters are used for variables, observations and estimated quantities. In their estimated versions the parameter matrices β and Γ thus are written B and C. Sometimes we will find it convenient to use symbols as $\hat{R}$ and $\hat{\Psi}$ for the estimated correspondents to R and Ψ. The column vectors of the endogenous and the predetermined variables at time point t are denoted y_t and z_t respectively. Observation matrices are written with capitals. Y is the matrix $(y_1, y_2, .. , y_T)'$ and Z the matrix $(z_1,z_2,..,z_T)'$. When the observations on the endogenous variables are required in vector form we will write vecp Y for the pT x 1 vector $(Y'_1,.., Y'_p)'$ and vecn Y for the nT x 1 vector $(Y'_1, . . ,Y'_n)'$. Y_i is the T x 1 vector containing the observations on the i^{th} member of y_t. Disturbances are denoted e.g. by δ_t and ϵ_t; the corresponding column vectors of "observations" are written δ and ϵ. The covariance matrices of δ_t and ϵ_t, which are of order p, are denoted Ψ and Ξ ; the covariance matrices of δ and ϵ are written Ψ_L and Ξ_L respectively. The meaning of $A \otimes B$, the Kronecker product of two matrices A and B, is the matrix whose ij^{th} submatrix is $a_{ij}B$, where a_{ij} is the ij^{th} element of A. The identity matrix of order T, say, is normally denoted I_T.

References to formulae in the text are given within parentheses. A reference to formula 2 in section 3 of chapter 4 thus is written (4.3.2). Sections and subsections of the chapters are denoted e.g. 4.3 and 4.3.1 respectively.

CHAPTER 1: A GENERAL INTRODUCTION TO THE SUBJECT

1.1 Simultaneity and serial correlation

Two major complications in the estimation of simultaneous equation systems are (I) the correlations between the disturbance terms and some explanatory variables and (II) serially correlated disturbances. During the last three decades, ever since the important article by Haavelmo (1943), considerable efforts have been devoted to the construction of methods that deal with (I). The reason being the fact that the standard estimation method for a single regression equation, ordinary least squares, does not produce estimates with the desirable property of consistency. The problem of serially correlated disturbances (II) is important in the estimation of single equations where it has been extensively treated. For simultaneous equations, however, the attention has been focussed less on this problem than on (I).

A usual situation in empirical econometric research is that all the factors that disturb the estimation procedure and the model construction cannot be accounted for. The researcher may then have to decide what factors are most important, incorporate them in the analysis and ignore the rest. Referring to the two major complications mentioned earlier, a choice has usually been necessary, and characteristically it is the simultaneity that is dealt with in the estimation of macroeconomic models, while the autocorrelation is ignored. Implicit in the analysis is then a ranking of the simultaneity as a factor of greater importance. Whereas theoretically the new and specific feature of simultaneous equations is the simultaneity, this motivating the stress in the pure theoretical developments, the relevance in ranking simultaneity as more important in empirical econometrics is doubtful. Fair (1972A appendix B, 1973) obtains results that for small and perhaps even medium-sized models indicate that serial correlation problems are likely to be more severe than those of simultaneity. Goldfeld and Quandt (1972) find through Monte Carlo experiments that it may be more important to take account of the serial correlation. The generality of these conclusions is limited, referring to the one model estimated by Fair and the high importance of the degree of serial correlation as regards Goldfeld and Quandt, but the indication is nevertheless of importance.

As a conclusion, referring also to the frequency of autocorrelation in econometric models—see section 1.3—and the consequences as discussed in the following section, there is a need for estimation methods that account

for serial correlation. In the following we will discuss methods making a choice between either of the two factors simultaneity or serially correlated disturbances unnecessary, the estimators will account for both.

1.2 The consequences of serial correlation

To somewhat illustrate the consequences of serially correlated disturbances we start with the simple one-relational model

$$y_t = z_t\gamma + \delta_t \tag{1}$$

where y_t and δ_t are two stochastic variables, z_t a 1 x m vector of exogenous—usually assumed nonstochastic—variables and γ a m x 1 vector of unknown parameters. Writing the model for all T (t = 1, T) observations simultaneously

$$Y = Z\gamma + \delta \tag{2}$$

the ordinary least squares (OLS) estimator c of the parameter vector is obtained as

$$c = (Z'\,Z)^{-1}\,Z'Y, \tag{3}$$

which has desirable optimal properties under appropriate assumptions. Now if one of the standard assumptions, that of no serial correlation amongst the δ_t's, is relaxed we get the following properties. The estimator is (I) unbiased but (II) inefficient, the sampling variances are probably large compared to a different method of estimation, (III) the usual covariance matrix is biased and is likely to seriously underestimate the true variance and (IV) the predictions are inefficient. The properties are weakened further in the case where y_t lagged is used as explanatory variable. The estimator then is (V) biased, this holds even if the disturbances are not serially correlated, and (VI) inconsistent.

We will not show that the stated properties hold, this is performed in the textbooks—see e.g. Johnston (1972)—but only make some comments. The efficient estimator when the disturbances are serially correlated is the generalized least squares (GLS) method. Comparing GLS and OLS we get an indication of the efficiency loss. See the following example with stochastic exogenous variables where the asymptotic variances are compared.

Example 1. Specifying the model as $y_t = \gamma_1 z_t + \delta_t$ where $z_t = \theta z_{t-1} + w_t$, $\delta_t = \rho\delta_{t-1} + \delta_t^+$ and all the symbols denote scalars, we compare the asymptotic variances of the estimates of γ_1 as obtained by the OLS and GLS methods. Under the assumptions that the δ_t^+'s are distributed independently

$N(0, \sigma^2)$, that the w_t's are independent of the δ_t^+'s and serially uncorrelated and that the absolute values of the parameters ρ and θ are less than one, Rao and Griliches (1969) obtain the asymptotic formulae. In economic applications it is usual that both the observable variables and the disturbances have positive serial correlation. We therefore give the asymptotic efficiency of OLS, i.e. Var(GLS)/Var(OLS) for some selected positive values of ρ and θ:

ρ	θ	*Var*(GLS)/*Var*(OLS)
.9	.9	.105
.9	.1	.097
.9	.0	.105
.8	.8	.220
.1	.9	.996

As is seen the loss in efficiency is considerable when ρ is large.

While the properties of the estimation methods for single equations in the case of serial correlation are well explored[1], this is not so for the estimators of simultaneous equations. As regards large sample properties one can expect that consistency is preserved when the model does not contain lagged endogenous variables; this is for the k-class estimators indicated by Fisk (1967). That the loss of asymptotic efficiency can be considerable may be demonstrated from the asymptotic covariance matrices given in our development. Goldfeld and Quandt (1972) demonstrate in their Monte Carlo study that there also is a large impact in terms of efficiency loss for finite samples; this is confirmed in our chapter 7.

Amemiya (1966) extends the analysis of Watson (1955) and Watson and Hannan (1956), who calculated the lower bounds of efficiency for the regression estimator and the maximum and minimum bias of the estimator of the covariance matrix, to the simultaneous equation case. The superiority of a modification of the two stage least squares method over the usual is indicated.

1.3 The presence of autocorrelated disturbances in econometric models

The purpose of this section is to give an indication as to the frequency and type of serially correlated errors and to discuss their sources. The measure of

[1] Lyttkens (1964) gives the general asymptotic covariance matrix for the regression coefficients estimated by OLS allowing for serially correlated disturbances and random exogenous explanatory variables.

serial correlation usually attached to the estimated relations of econometric models is the Durbin-Watson "d"-statistic. In section 2.4 different tests for autocorrelation are discussed. The conclusion is, that the "d"-statistic is a useful indicator and that a calculated "d"-value, not only in the region where the hypothesis of no serial correlation is rejected, but also in the inconclusive region, indicates serial correlation. This conclusion is based on the usual patterns of economic variables. It is not valid for autoregressive relations.

In table 1 the frequencies of the calculated "d"-values in the regions of non-rejection, inconclusiveness and rejection are displayed for some constructed simultaneous equation systems. The frequencies are tabulated at the 1% and 5% significance levels and are ordered with the regions indicating negative serial correlation to the left and those indicating positive serial correlation to the right. The numbers are not exact; in some models one or two relations are estimated using a smaller, but not exactly specified, number of observations. In some models different versions of separate relations are given. The former obscurity affects the significance limits of the "d"-statistic and as regards the second it is not always clear which relation belongs to the final version. The doubtful cases are however very few and cannot affect the conclusions of the table.

It is immediately clear that positive autocorrelation is frequent in these models—especially in the quarterly ones. The presence of autocorrelation may even be more frequent than the numbers show, since the model-builders probably have used the "d"-statistic—among other criteria—in choosing good versions of relations and models. Furthermore; the strength of the serial correlation probably is underestimated in the relations that contain lagged endogenous variables. At least for Brooking's model the number of relations with serially correlated disturbances is larger than given in the table; 16 relations not included in the table were estimated with a given autocorrelation structure specified.

The models in table 1 are not chosen according to the strict principles of random sampling, but the result supports what seems to be a wide-spread opinion; autocorrelated disturbances are usual in econometric models and the problem is accentuated in quarterly models. The frequencies obtained are alone sufficient to motivate the consideration of more sophisticated estimation methods that account for serially correlated disturbances.

A few minor comments in connection to table 1 remain to be given, The OBE, Brooking and Chow-Moore models are given in Hickman (1972). The source of the Klein-Goldberger model is Klein and Goldberger (1955) and of the Wharton EFU model Evans and Klein (1968). The models of Argentina

TABLE 1. Frequencies of calculated values of the Durbin-Watson "d"-statistic tabulated at the 1% and the 5% significance levels. Notations: PA = critical region for Positive Autocorrelation, NA = critical region for Negative Autocorrelation, I = Inconclusive region, NR = region of NonRejection.

Model	5 % level					1 % level				
	NA	I	NR	I	PA	NA	I	NR	I	PA
OBE quarterly data 55 observations		1	7	5	37			12	5	33
Brooking quarterly data 46 observations	3		19	16	29	1	2	28	10	26
Chow-Moore quarterly data 78 observations		1	6	4	9			9	3	8
Wharton EFU quarterly data 68 observations			6	3	35			8	5	31
Klein-Goldberger yearly data 18 observations			5	7	2			9	3	2
UNCTAD-staff Argentina, yearly data 15 observations	1	2	19	32	8		2	44	12	4
UNCTAD-staff India, yearly data 20 observations		1	26	13	8			37	5	6

and India, all others concern USA, are discussed in Ball (1973). The data of the Chow-Moore model are four-quarter differences. The first order serial correlations were calculated for the relations of the OBE-model after a normalization procedure. 51 of the calculated 56 correlations were significant at the 5% level.

1.3.1 Sources of autocorrelated disturbances. The next subject is a short discussion of the sources of serially correlated disturbances in econometric models. A useful observation for a start is that positive autocorrelation in the original values of economic time series is widely prevalent. The principal reasons for this are economic growth and business cycles. The omission of an explanatory variable (I) in a separate relation is therefore likely to cause posi-

tive autocorrelation due to the serial correlation in the omitted variable, whose behavior is reflected in the disturbance term. The cause of this omission need not be pure misspecification; small models have many attractive features to modelbuilders, but since many variables are deliberately excluded, autocorrelated disturbances seem more likely to occur in smaller than in larger models. A second important source of serial correlation is misspecification of the functional form (II) of a separate relation. The specification of a linear relationship when the true functional form is nonlinear, may very well result in serially correlated errors. A third source is the official time series published. The smoothing process (III) by means of which the government statistician interpolate quarterly values from basic annual information, or make seasonal adjustments, or chain together successive samples may simply be the culprit. The fourth source is termed lagged effect of temporary shocks (IV) distributed over a number of periods. In many cases one can expect that factors such as catastrophes have effects over more than one time period. If the lagged effect is not dealt with in the model, positive serial correlation is probable. The next source is the time period of the observations (V). It seems to be a general opinion, that problems of serially correlated errors are more pronounced in models that are quarterly, than in models that are yearly. This opinion is supported by table 1, and is often motivated referring to the inertia of economic variables. The disturbances express the effect of factors that are not explicitely accounted for in the model and the development of these economic factors often show more regularity than a pure random series. I.e., the "erratic" element may persist longer than the duration of a particular observation period and induce positive serial correlation. This effect that should be more pervasive when the model is fitted to quarterly or monthly data, is however only one side of the matter and is upweighted by other factors. One important factor working in the opposite direction is the lag structure (VI) of the true economic relationships. Liu (1969) argues that monthly lags are likely to be prevalent and shows that the use of quarterly or annual data in this case may cause serial correlation. A seventh source is measurement errors (VII). Errors of measurement often affect successive observations in the same way and may therefore introduce serial correlation.[1] The last source we mention is serial correlation inherent in the model specification (VIII). An example of this is the Houthakker-Taylor dynamic demand model, see Dhrymes (1971 ch. 4:6), which is specified as

[1] The sources III and VII refer rather to the residuals than to the disturbances. The former may be serially correlated even if the latter are not.

$$y_t = a_0 + a_1 x_t + a_2 s_t + \delta_t \tag{1}$$

$$s_t - s_{t-1} = \beta_0 s_{t-1} + \beta_1 y_t \tag{2}$$

with the variables: y demand for the commodity in question, x aggregate income and s habit persistence, which might be interpreted as the psychological stock of the commodity possessed by the individuals. Reducing the model to one relation we get

$$(1 - \beta_1 a_2) y_t = -a_0 \beta_0 + a_1 - a_1 (1 - \beta_0) x_{t-1} + (1 - \beta_0) y_{t-1} + \\ + \delta_t - (1 - \beta_0) \delta_{t-1} \tag{3}$$

an equation in observable quantities with disturbances generated according to a first order moving average process under usual assumptions about δ_t. In this connection we also mention the relation between the form of the geometric distributed lag model and the presence of serial correlation in the ordinary regression model, see Dhrymes (1971).

An important reason for our treatment of sources of serial correlation is that the solution to be adopted in a specific case depends on the source. Thus if the source is misspecification the appropriate action is to include the lacking explanatory variable or change the functional form of the model and so forth. A possible way to discover misspecification due to omitted variables is to regress the residuals on variables which on *a priori* grounds might be considered relevant as explanatory. The residuals can also give a rough indication as to the functional form of the regression equation. Autocorrelation inherent in the model specification is dealt with using the disturbance structure specified in the estimation.

It is only when the flaws in the model have been investigated and lacking important factors have been explicitely introduced, that the estimation methods suggested in the following chapters should be considered. A further requirement to use the methods is the specification of a reasonable correlation structure for the disturbances.

1.4 The specification of a disturbance pattern

The need for more sophisticated estimation methods accounting for serial correlation is clearly manifested by the preceeding sections. It is further supported by what seems to be the current practice of most model proprietors who issue regular forecasts. They adjust the forecasts by changing the values

of the constant terms of the relations and by adjusting the values of the exogenous variables used in the forecasts. Probably the subjective fine-tuning of forecasts never can be dispensed with, due to the difficulties of reproducing the real world in an econometric model, but the very use of econometric tools is an attempt to treat some of the problems of inter-subjectivity. It is therefore vital to try to further diminish the need for different kinds of "adjustments".

Fair (1972A, 1974) who has constructed a rather small model of the United States' economy is successful in producing forecasts without constant-term adjustments. The characteristic feature of his estimation procedure is the specification and use of a process generating the disturbances. The results show that the development of more efficient estimation techniques is a promising line of development.

One important component and the weakness in the construction of the outlined sophisticated estimation methods is the specification of the variation pattern of the disturbances. With a wrong specification the properties of the resulting estimates will not be optimal. In practice the two needs that should be met in the specification is realism, regarding the phenomena studied, and, as regards estimation, simplicity. It seems to be a wide-spread opinion that a first order autoregressive process satisfactorily can represent the serial correlation of economic models; at least for annual and for certain quarterly time series, cf. Goldberger (1965, p. 236) and Malinvaud (1970, p. 531). This kind of process, which is relatively simple to handle, is consequently dominating in the attempts made to handle serial correlation -as is obvious from section 1.5 -and is the one that is to be used in this study. The use of the first-order autoregressive process is also natural considering the standard test for serial correlation, i.e. the Durbin-Watson "d"-statistic, see section 2.4. This test is powerful to the alternative hypothesis of a first order autoregressive process. Furthermore the first-order process is, for many estimation methods, easily extended to higher order processes. The estimated coefficients can give useful information as regards the dynamic specification of the model, but this is more a problem of modelbuilding, a subject which will not be treated here.

1.5 Some suggested estimators

In this section we will recapitulate some important developments concerning estimators in the case of serially correlated disturbances. The main purpose of suggesting new or modifying the classical estimators, is to gain in efficiency

or–in the more severe situations–to get consistent estimates. To obtain estimates that are more efficient, some information concerning the type of serial correlation is required. The general approach has been to assume that the disturbances δ_t of the system[1]

$$y_t = \beta y_t + \Gamma z_t + \delta_t \tag{1}$$

are generated according to an autoregressive process

$$\delta_t = R\,\delta_{t-1} + \delta_t^+ \tag{2}$$

which sometimes is specified as being of higher order than the first written here. Some researchers have considered other processes; so has Phillips (1966) studied maximum likelihood methods in the moving average process case. But the overwhelming stress is on autoregressive processes and it is these that will be taken up in the following.

The case where R is specified as diagonal in (2) deserves special attention. As is seen in subsection 5.3.1 it is only in this limited-information situation that estimation methods utilizing only the specification of the separate relation to be estimated, i.e. limited-information methods, are strictly feasible. Sargan (1961), who is the first to consider estimators when a specified process is assumed to generate the disturbances, argues that there is little point in assuming that the autoregression structure matrix R is diagonal unless the covariance matrix of δ_t^+ also is assumed diagonal. And as regards the limited-information methods the off-diagonal elements of this covariance matrix are not accounted for. The stratification of the methods that are to be presented is: Maximum Likelihood-(ML), Two Stage Least Squares-(2SLS) and other methods.

1.5.1 ML-methods. The Full-Information Maximum Likelihood (FIML) method is discussed in chapter 3 and therefore the treatment here is kept short. The approach is, as for all methods, based on the transformed system obtained by inserting (1) into (2)

$$y_t = Ry_{t-1} + \beta y_t - R\beta y_{t-1} + \Gamma z_t - R\Gamma z_{t-1} + \delta_t^+ \tag{3}$$

The distributional assumptions are placed on δ_t^+ and the corresponding likelihood function is used to derive the estimator. Sargan (1961) gave the general

[1] y_t is the p x 1 vector of endogenous variables, z_t the m x 1 vector of predetermined variables. β, Γ and R are parameter matrices. δ_t and δ_t^+ are of order p x 1. To simplify this summary the system is assumed to contain behavioral relations only.

definition of the FIML estimator, or the FIMLA estimator as it will be called in chapter 3, and provided the numerical solution to a limited-information estimator with some information concerning the form of the equations ignored. Later Hendry (1971, 1974) considered the estimator with R full. He applied a numerical method based on direct maximization from function values only, and provided an expression for the asymptotic covariance matrix. The estimator is derived in such a way that restrictions in the form of prescribed zero elements in R cannot be handled in practice. Other numerical methods have been considered by Goldfeld and Quandt (1972), in the case where R is diagonal and by Chow and Fair (1973). The treatment of the two latter is rather general. It allows easily for the extension to autoregressive processes of higher order than the first and the most important parameter restrictions can be handled.

1.5.2 2SLS-methods. The two stage least squares approaches presented here are modifications of the 2SLS-method of Basmann (1957) and Theil (1958). All except one are based on the transformed system (3) and require that R is diagonal.

The historical and established view of 2SLS as a double regression method means that the estimates of the parameters of the i^{th} relation[1]

$$Y_i = Y_{(i)}\beta'_{(i)} + Z_{(i)}\gamma'_{(i)} + \delta_i \tag{4}$$

are obtained as

$$(b_{(i)}c_{(i)})' = [(Y^*_{(i)}Z_{(i)})' (Y^*_{(i)}Z_{(i)})]^{-1}(Y^*_{(i)}Z_{(i)})'Y_i \tag{5}$$

where the $Y^*_{(i)}$'s are the predicted Y-values obtained from a first stage where the Y's are regressed on all the Z's;

$$Y^*_i = Z(Z'Z)^{-1}Z'Y_i\,. \tag{6}$$

The notations are exactly explained in section 2.1 and the derivation of the method is systematic in chapter 4.

This interpretation of 2SLS is the dominating as regards the modifications for the autoregressive case. The exception is the early effort by Sargan (1964). He defines an Instrumental Variables (IV) method which according to the IV interpretation of 2SLS is asymptotically equivalent to that method and to

[1] Y_i is a T x 1 vector of observations. $\beta_{(i)}$ and $\gamma_{(i)}$ of order 1 x n_i and 1 x m_i respectively denote the non-zero coefficients of relation i. $Y_{(i)}$ and $Z_{(i)}$ are the corresponding observation matrices.

Sargan's (1961) limited-information maximum likelihood method; both as applied to the transformed system.

Theil has suggested what he calls the generalized 2SLS method, see Theil (1958). The development corresponds to the generalization of the ordinary least squares method to the generalized least squares method. Writing the covariance matrix of δ_i as ψ_{Lii}, which is of order T, the second stage (5) is modified to

$$(b_{(i)}c_{(i)})' = [(Y^*_{(i)}Z_{(i)})' \psi^{-1}_{Lii} (Y^*_{(i)}Z_{(i)})]^{-1}(Y^*_{(i)}Z_{(i)})' \psi^{-1}_{Lii}Y_i \qquad (7)$$

With this definition the process generating the disturbances need not be of the form (2). However the first stage, where the Y^*_i's are obtained, means estimation of the reduced form where the disturbances are different from δ_i. Applying the generalized least squares principle here too Theil is forced to specify

$$Y^*_i = Z(Z'\psi^{-1}_{Lii}Z)^{-1} Z'\psi^{-1}_{Lii} Y_i . \qquad (8)$$

This stage is not valid unless the covariance matrix of the reduced form disturbances is "proportional" to the covariance matrix of the structural form disturbances ψ_{Lii}, which is a rather heavy restriction. The condition is satisfied e.g. when the disturbance vectors of each pair of relations have a covariance matrix equal to ψ_{Lii} except for a scalar. In the case of an autoregressive process where R is diagonal a sufficient condition is that all elements on the main diagonal are identical.

The first 2SLS estimator of Amemiya (1966) is S2SLS, a two stage least squares analogue of Sargans limited-information estimator above, Sargan (1961). Writing the i^{th} relation of the transformed system (3) with R diagonal as

$$Y_i = \rho_{ii}Y_{il} + Y_{(i)}\beta'_{(i)} - \rho_{ii}Y_{(i)l}\beta'_{(i)} + Z_{(i)}\gamma'_{(i)} - \rho_{ii}Z_{(i)l}\gamma'_{(i)} + \delta^+_i \qquad (9)$$

where $Y_{(i)l}$ is $Y_{(i)}$ lagged one period, more compactly as

$$Y^T_i = Y^T_{(i)}\beta'_{(i)} + Z^T_{(i)}\gamma'_{(i)} + \delta^+_i \qquad (10)$$

we have the basic equation for the S2SLS method. In the first stage $Y^T_{(i)}$ is purged by regressing the corresponding transformed variables on the predetermined, the lagged predetermined and the lagged endogenous variables of the original system (1). Then the second stage is applied to (10) with $Y^T_{(i)}$ replaced by the predicted values. In practice the ρ_{ii} parameter used in the transformation is unknown and has to be estimated. The S2SLS-estimator of ρ_{ii} is given by Amemiya as

$$r_{ii} = (d'_{il}d_{il})^{-1} d'_{il}d_i \qquad (11)$$

where d_i contains the residuals of the i^{th} equation of (1) as obtained using the S2SLS estimator of $\beta_{(i)}$ and $\gamma_{(i)}$. Estimates are obtained iteratively. The S2SLS estimator is consistent regardless whether ρ_{ii} is known or estimated and the asymptotic distributions are identical in the two cases.

The consequence of a wrong specification of the pattern of disturbances, e.g. when some fixed not true value of ρ_{ii} is chosen for the transformation (10), is in general that the S2SLS estimator is not consistent. Amemiya (1966) therefore develops the modified S2SLS method (MS2SLS) where only the predetermined and the lagged predetermined variables and not the lagged endogenous variables are used in the first stage. The MS2SLS estimator is consistent in more general circumstances but the price that is paid is a loss in asymptotic efficiency. It should also be noted that the MS2SLS method requires that all the predetermined variables are purely exogenous.

A main difficulty in the application of the S2SLS method is the large number of regressors in the first stage. The number is even for medium-sized models likely to exceed the number of observations available. Fair (1970) has tried to meet this drawback. One method that he suggests is similar to S2SLS but uses knowledge of the reduced form. Solving for the reduced form of the transformed system it may well be that some variables say z_{1t} and z_{1t-1} do not enter as separate variables, but only as $z_{1t} - \rho_{jj} z_{1t-1}$ where ρ_{jj} is an element in the diagonal matrix R. By utilization of such restrictions the number of regressors in the first stage is decreased. The asymptotic covariance matrix of this method is identical to that of the S2SLS method also when the elements of R are consistently estimated. Fair indicates that the fact that S2SLS uses, so to say, unnecessary instruments implies increased bias. To further decrease the number of regressors in the first stage Fair (1970) suggests the X2SLS estimator. The correct application of this method is based on the assumption that $R = \rho_0 I_p$ where I_p is the identity matrix of order p and ρ_0 is a scalar. Estimating (9) the regressors of the first stage are Y_{il}, $Y_{(i)l}$, $Z_{(i)}$, $Z_{(i)l}$ and $Z_{(j)} - \rho_0 Z_{(j)l}$ where $Z_{(j)}$ contains the predetermined variables that are not included in $Z_{(i)}$. The X2SLS estimator is asymptotically less efficient than the earlier methods but due to the smaller number of regressors where $Z_{(j)} - \rho_0 Z_{(j)l}$ is used instead of $Z_{(j)}$ and $Z_{(j)l}$ the small sample properties may be better.

1.5.3 Other methods. Remaining to discuss are Instrumental Variables (IV) methods more generally and certain full information methods. As already has been mentioned 2SLS may be interpreted as an instrumental variables method; the reason why a certain section is devoted to two stage least squares methods

is the numerous references. The IV interpretation is a base for developing methods that are possible to apply when the sample is undersized, and 2SLS consequently breaks down in the first stage, methods with desirable asymptotic properties.

Handling the full-information case, where R need not be diagonal, first we rewrite the i^{th} relation, cf. (4), as

$$Y_i = X_i a_i + \delta_i \tag{12}$$

where $a_i = (\beta'_{(i)}\gamma'_{(i)})$ and $X_i = (Y_{(i)}Z_{(i)})$. The whole system with the p relations is written as

$$\text{vecp } Y = X\,a + \delta\ , \tag{13}$$

vecp $Y = (Y'_1, ..., Y'_p)'$, $a' = (a'_1, ..., a'_p)$, $\delta' = (\delta'_1, ..., \delta'_p)$ and X is a block-diagonal matrix containing the X_i's. The transformed system is now obtained as[2]

$$\text{vecp } Y - (R \otimes I_T)\,\text{vecp } Y_l = (X - (R \otimes I_T)X_l)\,a + \delta^+ \tag{14}$$

the difference in comparison to (3) being that observation vectors now are used. The instruments which we will consider are based on predicted Y-values obtained from the restricted reduced form and not the unrestricted reduced form as is the case for 2SLS. The methods are therefore structural form oriented, cf. chapter 4, using estimates of the structural form parameters. This IV approach has been suggested for InterDependent (ID) systems without serially correlated disturbances by Lyttkens, see e.g. Mosbaek and Wold (1970), and by Brundy and Jorgenson (1971).[3] The development for the transformed system (14) is due to Fair (1972B) who obtains the estimator

$$\begin{aligned} a = {} & [(X^* - (\hat{R} \otimes I_T)X_l)'\,(\hat{\Omega}^{-1} \otimes I_T)(X - (\hat{R} \otimes I_T)X_l)]^{-1} \\ & (X^* - (\hat{R} \otimes I_T)X_l)'(\hat{\Omega}^{-1} \otimes I_T)(\text{vecp } Y - (\hat{R} \otimes I_T)\,\text{vecp } Y_l) \end{aligned} \tag{15}$$

where $\hat{\Omega}$ and $\hat{R}$ are consistent estimates of the covariance matrix of δ^+ and the autoregression structure matrix respectively. X^* is X with the values of the current endogenous variables replaced by predicted values. The estimates $\hat{\Omega}$ and $\hat{R}$ are obtained from residuals cf. (11); R can be estimated through a Zellner "seemingly unrelated regressions" approach i.e. a generalization of the formula (11). Initial consistent estimates of the structural parameters can be obtained e.g. by treating all lagged endogenous variables of the original

[2] The Kronecker product $R \otimes I_T$ is a pT x pT matrix whose ij^{th} submatrix is $\rho_{ij} I_T$.

[3] Bergström (1974) has applied Lyttkens' Iterative Instrumental Variables method to a number of models.

system (1) as endogenous and estimating that system by an IV method. Iteration between estimates of R and estimates of α produces, provided convergence is obtained and given regularity conditions are satisfied, an estimator with the same asymptotic properties as the FIMLA-estimator. A closer examination of estimators similar to (15) will be performed in chapter 5.

The estimator for the limited-information case, where R is diagonal, is obtained by Fair (1972B) analogously to the above derivation. An estimating equation is specified for every relation as

$$a_i = [(X_i^* - \hat{\rho}_{ii}X_{il})'(X_i - \hat{\rho}_{ii}X_{il})]^{-1}(X_i^* - \hat{\rho}_{ii}X_{il})'(Y_i - \hat{\rho}_{ii}Y_{il}). \quad (16)$$

The consistent estimate $\hat{\rho}_{ii}$ is obtained through a regression similar to that given in (11). Asymptotically the estimator is equivalent to Amemiyas S2SLS method.

Fair also makes a suggestion for the limited-information estimation of models with a nondiagonal R. Assume, to illustrate, that the i^{th} row contains a nonzero element ρ_{ij} ($i \neq j$). Then the following transformed equation is used

$$Y_i - \rho_{ij}\delta_{jl} = \rho_{ii}Y_{il} + X_i a_i - \rho_{ii}X_{il}a_i + \delta_i^+, \quad (17)$$

which is identical to (9) except for the term $\rho_{ij}\delta_{jl}$ on the left hand side. Provided consistent estimates of ρ_{ij} and δ_{jl} are available the estimation can proceed as for the limited-information case.

A late contribution is made by Dhrymes, Berner and Cummins (1974) for the case where R is diagonal. They discuss a two stage least squares type method, cf. our section 5.3.2, where Y in the first step is regressed on Z and Z_l. The predicted Y-values are utilized obtaining consistent parameter estimates and associated residuals. Through ordinary least squares as applied to the latter consistent estimates of the elements of R are obtained. New estimates of the structural parameters of the transformed system (14) are calculated in the following step. Utilizing the residuals R is re-estimated and new estimates of the structural parameters of (14) can be obtained. The procedure is repeated until convergence is reached. Dhrymes, Berner and Cummins also suggest an IV modification of their method by changing the very first step. Furthermore they consider a method where predicted Y-values are obtained through the final form. The values are used as "regressors" in a generalized least squares step; the resulting estimator is asymptotically less efficient than their two stage least squares type estimator.

The method defined through (15) is similar to the approach of the Three

Stage Least Squares (3SLS) estimator. That method and alternatives are discussed in chapter 5, but this is the place to at least mention some attempts. Dhrymes, see Dhrymes and Erlat (1974), has recently suggested a 3SLS estimator, with the asymptotic distribution of the system (14). He assumes that R does not contain any zero elements. A 3SLS approach when the structure of the covariance matrix of the disturbances is more general than implied by the autoregressive process (2) is mentioned by Madansky (1964). Wickens (1969) discusses this approach and some alternatives.

For the less general case of a system of vector regressions, or in other words the system (1) with an *a priori* zero β matrix, reference is made to Parks (1967) and Guilkey and Schmidt (1973).

CHAPTER 2: THE SPECIFICATION OF THE MODEL

2.1 The basic system

2.1.1 The classic interdependent system. The basic system considered in this monograph is the linear interdependent system written in the Structural Form (SF) as

$$y_t = \beta y_t + \Gamma z_t + \delta_t \qquad (1)$$

where $y_t = (y_{1t},..., y_{nt})'$ is the n x 1 vector of the endogenous observable variables, $z_t = (z_{1t},..., z_{mt})'$ the m x 1 vector of the predetermined, i.e. the lagged endogenous and the exogenous, observable variables and δ_t the n x 1 vector of the unobservable structural disturbances. $\beta = (\beta_{ij})$ and $\Gamma = (\gamma_{ij})$ are parameter matrices of order n x n and n x m respectively. The subscript t indicates that the structural form is written at timepoint t . The relations in (1) are of two different kinds; behavioral relations and identities. A behavioral relation is usually supposed to describe the behavior of a group of economic agents. We write it as

$$y_{it} = \sum_j \beta_{ij} y_{jt} + \sum_j \gamma_{ij} z_{jt} + \delta_{it} \qquad (2)$$

where the summations on the right hand side include the nonzero coefficients of the i^{th} row of β and Γ respectively. The identities have known parameters only and are not stochastic, i.e. are error free. The system (1) is reorded and with the use of partitioned matrices rewritten as

$$\begin{pmatrix} y_{Bt} \\ y_{It} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{Bt} \\ y_{It} \end{pmatrix} + \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} z_t + \begin{pmatrix} \delta_{Bt} \\ 0 \end{pmatrix} \qquad (3)$$

where the p behavioral relations are placed in the p first and the q = n–p identities in the q last equations. The orders of the parameter matrices are: β_{11} p x p, β_{12} p x q, Γ_1 p x m, β_{21} q x p, β_{22} q x q and Γ_2 q x m. The orders of the vectors of variables follow, however by the notation δ_t we will in the following sections 2.2–2.5 mean the p x 1 vector δ_{Bt} unless otherwise is stated.

Mostly we are going to work with a system asymmetrically specified as in (1) and (3) with the main diagonal of β assumed to contain zeroes only. Provided the matrix $(I - \beta)$ is nonsingular, the Reduced Form (RF) is defined as

$$y_t = (I - \beta)^{-1}\Gamma z_t + (I - \beta)^{-1}\delta_t = \Phi z_t + \epsilon_t \qquad (4)$$

where $\Phi = (\varphi_{ij})$ of order n x m denotes the reduced form parameters and $\epsilon_t = (\epsilon_{1t},..., \epsilon_{nt})'$ the reduced form disturbances.

Rewriting (1) and (4) as

$$\text{SF:} \quad y_t = \beta y_t + \Gamma z_t + \delta_t \tag{5}$$

$$\text{RF:} \quad y_t = \Phi z_t + \epsilon_t \tag{6}$$

we define what we will call a *CLassic InterDependent (CLID) system.* Thorough presentations of CLID systems are found in e.g. Christ (1966), Fisk (1967), Dhrymes (1970), Malinvaud (1970), Schönfeld (1971A). Theil (1971) and Johnston (1972).

2.1.2 The reformulated interdependent system. A large part of our developments will be based on the results applying a reformulation principle. Reformulating the CLID system (5) and (6) we proceed as follows. Defining the n x 1 vector η_t^* as the systematic part of the reduced form

$$\eta_t^* = \Phi z_t \; ; \; y_t = \eta_t^* + \epsilon_t \tag{7, 8}$$

the structural form (5) can be respecified to

$$y_t = \beta \eta_t^* + \Gamma z_t + \epsilon_t \tag{9}$$

by inserting (8) in the right hand side of (5) and utilizing the relation $\beta\epsilon_t + \delta_t = \epsilon_t$ which follows from (4). Formula (9) defines the structural form of the *REformulated InterDependent (REID) system.* The reduced form of the REID system is identical to the reduced form of the CLID system. Thus the REID system is written as

$$\text{SF:} \quad y_t = \beta \eta_t^* + \Gamma z_t + \epsilon_t \tag{10}$$

$$\text{RF:} \quad y_t = \Phi z_t + \epsilon_t \tag{11}$$

It is seen that the disturbances of the reduced form and of the structural form coincide.

We have derived the REID system starting from the CLID system but it is easily realized that we can go the other way round; i.e. the systems are compatible in the sense that the one implies the other. The fundamental difference between the two is that the REID system but not the CLID system admits a specification of both the reduced and structural form as unbiased predictors. The REID system offers a predictive interpretation of the structural form. The specification of η_t^* as the systematic part of the endogenous variables or

$$\eta_{it}^* = E(y_{it}|z_t), \quad (i=1,n) \tag{12}$$

admits the predictor specification

$$E(y_{it}|\eta^*_{(i)t}, z_{(i)t}) = \beta_{(i)}\eta^*_{(i)t} + \gamma_{(i)}z_{(i)t}, \quad (i=1,n) \tag{13}$$

where $\eta^*_{(i)t}$ and $z_{(i)t}$ denote the components of η^*_t and z_t occurring in the i^{th} relation and $\beta_{(i)}$ and $\gamma_{(i)}$ the i^{th} rows of β and Γ respectively with the prescribed zero elements eliminated. The conditional expectations (12) define the *REID specification.* This specification applies as we consider it to all equations of the system, identity as well as behavioral relation. For some models it may be inconvenient to specify the identities asymmetrically with one endogenous variable on the left hand side. An alternative suggested by Lyttkens, see Mosbaek and Wold (1970, ch. 11), is the partly respecified system. In this approach only the behavioral relations are respecified while the identities are kept in the original form.

The result (13) is in contrast to the corresponding conditional expectations for the CLID system where the predictor specification of the reduced form generally does not admit a specification of the structural form as an unbiased predictor i.e.

$$E(y_{it}|y_{(i)t}, z_{(i)t}) \neq \beta_{(i)}y_{(i)t} + \gamma_{(i)}z_{(i)t}, \ (i=1,p) \tag{14}$$

Consequences of (12) are that the expected values of the disturbances ϵ_t are zero and that these disturbances are uncorrelated with all the predetermined variables.[1] The specification (12) however, does not imply that the structural form disturbances of the CLID system δ_t are uncorrelated with the endogenous variables, which would be illogical and therefore the inequalities in (14) are obtained.

The REID system was developed by Wold (1965) while the REID specification was introduced by Ågren and Wold (1969); recent descriptions of the approach are found in Wold and Lyttkens (1969), Mosbaek and Wold (1970) and Lyttkens (1973).

2.1.3 Basic assumptions. An important problem considering interdependent systems is to obtain estimates of the parameters of the behavioral relations. To succeed with this not only a sample of observations on the endogenous and predetermined variables is required but also assumptions concerning the system and the behavior of the disturbances and the predetermined variables. Some of these are required for the very existence of estimates whilst other are made to obtain estimates with desirable properties. Assuming that the y's and the z's

[1] The original REID specification was made for the case of stochastic exogenous variables.

are observable for T periods (t=1,T) we define the following observation matrices

$$Y = (y_1,...,y_T)' \ , \ Z = (z_1,...,z_T)'. \qquad (15)$$

The assumptions that follow are mainly based on those by Christ (1966, pp. 350–357).

Assumption 1: The system is linear in the parameters and the variables.
Assumption 2: The matrix $(I - \beta)$ is nonsingular.
Assumption 3: The predetermined variables are not linearly dependent.
Assumption 4: The system is identified.
Assumption 5: The values of the lagged endogenous variables $y_{t-1},..., y_{t-h}$, where h denotes the maximal lag, are known for t=1.
Assumption 6: The moment matrix of the predetermined variables is well-behaved in the limit i.e.

$$\text{plim } T^{-1}Z'Z = \lim E(T^{-1}Z'Z) = M_{ZZ}$$

which is nonstochastic, has finite elements and is nonsingular.
Assumption 7: The exogenous variables z_t^e are bounded and conceptually held fixed in repeated samples. The limiting matrix $\lim_{T\to\infty} M^\nu$ is assumed to exist for $\nu = 1,2,.$ where

$$M^\nu = (T-\nu)^{-1} \sum_{t=1}^{T-\nu} z_t^e z_{t+\nu}^{e'} .$$

Assumption 8: The parameters of the components of z that are lagged endogenous are such that the system is stable.
Assumption 9: The random variables δ_t are stochastically independent vectors that are identically distributed with $E(\delta_t) = 0$ and $E(\delta_t\delta_t') = \Psi$, (t=1,T). The matrix Ψ is nonstochastic, has finite elements and is positive definite.
Assumption 10: The expectation of δ_t conditional on the exogenous variables at all time-points and the lagged endogenous variables at time-points 1,2,..., t, is identical to the zero vector.
Assumption 11: The disturbances have finite moments up to the fourth order.
Assumption 12: The disturbances are jointly normally distributed.

In the developments we also will need more specific assumptions than 1–12 above. These are invoked when required. The assumptions concerning the disturbance terms are in our list placed on the SF disturbances of the CLID system–δ_t. They do though, as is obvious from (4), in all essential aspects carry over to ϵ_t. The assumption of normally distributed disturbances is required for a discussion concerning asymptotic efficiency but is in general not necessary establishing asymptotic normality. Assumption 11 is only used showing asymptotic normality. Note that assumption 10 is closely connected with the

specification of the reduced form as an unbiased predictor in the previous subsection.

The inferences that will follow are, as is obvious from assumption 7, conditional on the values of the exogenous variables. The choice between a conditional approach, where the model is utilized to derive from the distribution of δ given a realization z^{eo} the distribution of y given z^{eo}, and an unconditional approach, where the joint distribution of y and z^e is obtained from the joint distribution of δ and z^e, depends ultimately on the economic problem. The issue is whether the endogenous variables should be examined within the framework of the exogenous variables or independently of the values of these variables. The conditional approach is the traditional in econometric research. A reason for this is the simplified statistical analysis. However, often the two approaches are equivalent in the sense of producing identical asymptotic distributions of various estimators. Therefore we may expect that our results of chapters 4 and 5, chapter 3 is in this aspect more general, hold for the unconditional case. To prove this one could work according to the approach used by Dhrymes (1970), who however does not allow for lagged endogenous variables, or the approach used by Edgerton (1973).

2.1.4 The GEID specification. In the REID specification (12) we considered the expectations of the endogenous variables conditional on all the predetermined variables. In the *GEneral InterDependent (GEID) specification* the expectations are conditional on linear combinations of the predetermined variables. This specification, which was developed for the case of stochastic exogenous variables by Wold (1965, 1966), still produces unbiased predictors for REID systems but is more general in the sense that the number of zero correlations implied is smaller. Specifying the structural form (10) as a predictor

$$E(y_{it}|\eta^*_{(i)t}, z_{(i)t}) = \beta_{(i)}\eta^*_{(i)t} + \gamma_{(i)}z_{(i)t} \qquad (16)$$

cf. (13), we obtain for the reduced form

$$E(y_{it}|[\Phi z_t]_i) = \eta^*_{it}$$

where in both (16) and (17) (i=1,n). $[\Phi z_t]_i$ denotes the i^{th} row of Φz_t. An implication of (16) is that ϵ_{it} is uncorrelated with the components of η^*_t and z_t occurring in the i^{th} relation, i.e. $\eta^*_{(i)t}$ and $z_{(i)t}$, but may be correlated with the other components. We then generally obtain fewer zero correlations as compared to the REID specification where ϵ_{it} (i=1,n) is uncorrelated with all the predetermined variables.

In our developments we will hardly at all use the GEID specification. The reason for somewhat discussing it is that this specification is the origin of the REID specification and as regards serial correlation has an unexpected feature.

As a start, we make a dual definition of the reduced form. In this chapter we have so far only dealt with the reduced form as obtained from the structural form

$$y_t = (I - \beta)^{-1} \Gamma z_t + \epsilon_t = \Phi z_t + \epsilon_t \,. \tag{18}$$

This will be referred to as the *Restricted Reduced Form* (RRF). The *Unrestricted Reduced Form* (URF) on the other hand is defined as the unrestricted projection of y_t on z_t or

$$y_t = E(y_t|z_t) + \epsilon_t \,. \tag{19}$$

For classic interdependent systems and the REID specification it follows from (19) and (12) that the RRF and the URF, where we denote $E(y_t|z_t) = \Pi z_t$, coincide in terms of population quantities. This equivalence does not apply to the GEID specification; since the conditional expectation of ϵ_{it} given z_t here is different from zero (Lyttkens 1967). Assuming that y_t and z_t are jointly normal Bodin (1974) utilizes the relation

$$\epsilon_t = A\, z_t + \epsilon_t^* \tag{20}$$

where A is the n x m matrix of $E(\epsilon_t|z_t) = A\, z_t$ and ϵ_t^* is a vector of disturbances uncorrelated with the z_t's, to show that the occurrence of lagged endogenous or autocorrelated exogenous variables in general induces serial correlation in the ϵ_t's even though the ϵ_t^*'s are assumed non-autocorrelated. This means that the GEID specification exhibits a serial correlation that is not related to the origins of serial correlation discussed in chapter 1. As is stressed by Ågren (1974), however, this serial correlation will not affect the properties of the estimator designed for the specification, cf. also chapter 4.

The GEID specification has been criticized on several grounds. An important point is that the corresponding system is fundamentally underidentified in the two senses: *1* for a given structure, i.e. a system where fixed zero elements of β and Γ are specified, several parameter sets—if more than one—have the same likelihood (Lyttkens pp. 165–168 in Mosbaek-Wold 1970) and *2* likelihoods corresponding to different structures where the same Y and Z enter are identical (Ågren 1974). Therefore the system is not testable provided no further restrictions are imposed. It should be emphasized that these points apply to the case of normally distributed errors.

2.2 The transformed system

The main purpose of this monograph is to develop estimators designed for the case of serially correlated errors. More precisely we consider the situation where the vector of structural disturbances is adequately represented by a first order vector autoregressive process

$$\delta_t = R\,\delta_{t-1} + \delta_t^+ \quad . \tag{1}$$

R is a p x p matrix of parameters where the sum of the elements of each row is an absolute value smaller than one. Some elements may *a priori* be zero; the case where R is diagonal is important and will be studied in detail. δ_t^+ is a p x 1 vector of disturbances.

Assume to simplify the discussion that the original system (2.1.1), where δ_t is expressed in y_t, z_t and the structural parameters, consists of p behavioral relations only.[1] Inserting (2.1.1) in (1) above we obtain the *transformed system*

$$y_t = \beta y_t + R(I - \beta)y_{t-1} + \Gamma z_t - R\Gamma z_{t-1} + \delta_t^+ \tag{2}$$

which is a base for our developments. To simplify the notations it is assumed that the y's and z's are observable for (T + 1) periods (t=0,T) when discussing the transformed system, which then is valid for (t=1,T). The point with the transformed system is that all assumptions involving δ_t earlier now can be transferred to δ_t^+. The price that is paid is the more complicated parameter structure and the presence of lagged endogenous variables even if z_t is purely exogenous. Before discussing a reformulated system we will make some modifications of the assumptions of subsection 2.1.3.

We assume that the main part of the assumptions made applies to (2) with its parameter structure and predetermined variables. The explicit modifications concern assumptions 1, 6, 9–12. The system (2) is not linear in the parameters and therefore assumption 1 is not valid. The nonlinear parameter structure requires some mild regularity conditions, cf. Malinvaud (1970 ch. 9.5), which we will not discuss. The assumption of a well-behaved matrix of moments in the limit, assumption 6, is extended to the moment matrix of the predetermined, lagged predetermined and lagged endogenous variables of the original system (2.1.1). Naturally the duplicate variables, say y_{jt-1} which is predetermined in the original system but also is introduced through the transformation, occur only once. The assumptions 9–12 now concern δ_t^+; the corresponding covariance matrix $E(\delta_t^+\delta_t^{+\prime})$ is denoted Ω.

[1] Identities are easily accomodated at the cost of notational complications mainly.

The specification of the process (1) for δ_t is equivalent to the specification

$$\epsilon_t = (I - \beta)^{-1}R(I - \beta)\epsilon_{t-1} + \epsilon_t^+ \tag{3}$$

or

$$\epsilon_t = R_\epsilon \epsilon_{t-1} + \epsilon_t^+ \tag{4}$$

for the reduced form disturbances. We could work with R_ϵ instead of R in the following, but due to the current practice, see section 1.5, where the zero-restrictions in R are important, the choice is R. Otherwise these restrictions would be difficult to handle.

2.2.1 The reformulated transformed system. Following the reformulation principle outlined in section 2.1.2 we will respecify the transformed system (2) to admit a predictor specification of the structural form. Defining the p x 1 vector η_t^{**} as the systematic part of the restricted reduced form of (2)

$$\eta_t^{**} = (I - \beta)^{-1}\Gamma z_t + (I - \beta)^{-1}R(I - \beta)y_{t-1} \cdot (I - \beta)^{-1}R\Gamma z_{t-1} \tag{5}$$

The system (2) can be respecified to

$$y_t = \beta\eta_t^{**} + R(I - \beta)y_{t-1} + \Gamma z_t - R\Gamma z_{t-1} + \epsilon_t^+ \tag{6}$$

Thus we have the structural form of the *REformulated InterDependent Autocorrelated disturbances* (REIDA) *system.* The reduced form of the REIDA system is identical to the reduced form of the transformed system or

$$y_t = \Phi z_t + (I - \beta)^{-1}R(I - \beta)y_{t-1} - (I - \beta)^{-1}R\Gamma z_{t-1} + \epsilon_t^+ \tag{7}$$

The definition of η_t^{**} in (5) and the REIDA specification

$$\eta_{it}^{**} = E(y_{it}|z_t, y_{t-1}, z_{t-1}), \tag{8}$$

where (i=1,p), admits that

$$\begin{aligned} E(y_{it}|\eta_{(i)t}^{**}, z_{(i)t}, y_{(i)t-1}, z_{(i)t-1}) &= \beta_{(i)}\eta_{(i)t}^{**} + \\ &+ [R(I - \beta)]_{(i)}y_{(i)t-1} + \gamma_{(i)}z_{(i)t} - [R\Gamma]_{(i)}z_{(i)t-1}, \end{aligned} \tag{9}$$

i.e. the expectation of the left hand side variable of relation i (i=1,p) given the components of η_t^{**}, z_t, y_{t-1} and z_{t-1} occurring on the right hand side, is an unbiased predictor. For the case of identities cf. subsection 2.1.2.

In accordance with the GEID specification we can decrease the number of variables upon which the y_t's are conditioned and still obtain unbiased pre-

dictors. Details of this are discussed in Selén (1973A, B) where also the implications of (7) and (8), these are parallels to the implications discussed in section 2.1.2, are explained.

A second possible approach to the reformulation in the case of serially correlated disturbances is to start from the REID system

$$y_t = B' \eta_t^* + \Gamma z_t + \epsilon_t \tag{10}$$

where the disturbances are assumed to be generated according to

$$\epsilon_t = R_\epsilon \epsilon_{t-1} + \epsilon_t^+ \tag{11}$$

cf. (4). Rewriting (11) utilizing (10) we get

$$y_t = B\eta_t^* + R_\epsilon y_{t-1} - R_\epsilon B\eta_{t-1}^* + \Gamma z_t - R_\epsilon \Gamma z_{t-1} + \epsilon_t^+ \tag{12}$$

To limit the discussion in the following we will make a choice between the two alternatives; the REIDA-system as written in (6) and the REID-system in (12). In principle this choice should be determined by considerations whether the CLID-system (2.1.1) and (2.2.2), which is the base for REIDA, or the REID system (10) is the primary system. In other words: Does the modelbuilder constructing a model and/or a behavioral relation, think in terms of conditional expectations, thus favouring a REID-system, or not? In chapter 4, where some estimators in the case of serially uncorrelated disturbances are discussed, the point of departure is the REID-system (10). A logical approach is to continue according to this line of development and use (12) in chapter 5, where some estimators designed for serially correlated disturbances are discussed. This will however not be done. We will there work according to CLID-REIDA. A reason is the possibilities of comparisons to other developments made, cf. section 1.5, where generally the autoregressive process is specified for δ_t and not ϵ_t, the base for (12). It is also simple to work with the system (6) interpreting ϵ_t^+ as the disturbance of a "new" system, differing from the original mainly through restrictions placed on the structural parameters. A similar interpretation of the disturbance ϵ_t^+ is for the alternative (10) and (12) not feasible. The implications of the REID specification concern ϵ_t and not ϵ_t^+; it is in fact easy to show, see Selén (1973A), that in general the disturbances ϵ_t^+ and the predetermined variables that are lagged endogenous are not uncorrelated. This means difficulties to obtain estimators that in the presence of lagged endogenous variables are consistent. The small sample properties of one estimator based on (10) is in the absence of lagged endogenous variables studied in Selén (1973B).

2.3 The identification problem: Some possible approaches

The identification problem concerns the specification of the model, i.e. the question whether the model is restrictive enough so that, given sufficiently large samples, the values of the parameters can be uniquely determined. A more formal statement is: Assuming that the parameters of the reduced form are known can we then obtain some or all of the values of the structural parameters uniquely? As is well known the system

$$y_t = B\, y_t + \Gamma\, z_t + \delta_t \tag{1}$$

is identified if and only if no equation can be expressed as a linear combination of the other equations of the model.

The crucial case when the δ_t's are autocorrelated is the presence of lagged endogenous variables in the vector z_t. The correlations between the current and lagged disturbances imply that the current disturbances and the lagged endogenous variables are correlated which means that the lagged endogenous variables cannot be taken as predetermined. The consequences of this is clearly seen by the following example used by Koopmans, Rubin and Leipnik (1950) and Fisher (1966). In the two-relational model

$$y_{1t} = \gamma_{11} y_{1t-1} + \gamma_{12} y_{2t-1} + \delta_{1t} \tag{2}$$

$$y_{2t} = \beta_{21} y_{1t} + \delta_{2t} \tag{3}$$

both equations are identified by the standard rank condition if the lagged endogenous variables are taken as predetermined. The following equation is obtained by lagging (3), multiplying it by a scalar γ and adding it to the first equation

$$y_{1t} = \gamma^*_{11} y_{1t-1} + \gamma^*_{12} y_{2t-1} + \delta^*_{1t} \tag{4}$$

where

$$\gamma^*_{11} = \gamma_{11} + \lambda\beta_{21},\ \gamma^*_{12} = \gamma_{12} - \lambda,\ \delta^*_{1t} = \delta_{1t} + \lambda\delta_{2t-1}$$

Generally, in the case where the autocorrelation properties of the disturbances are unspecified, δ^*_{1t} and δ_{1t} are indistinguishable which means that (2) and (4) are also indistinguishable. With restrictions on the autocorrelation pattern the situation is different. If the disturbances are serially uncorrelated, δ^*_{1t} and δ_{1t} are distinguishable since the former but not the latter is correlated with the lagged endogenous variables and (2) is therefore identified.

The trouble in the case of serial correlation is thus that transformations in time may become admissible i.e. that observational information cannot

distinguish the i^{th} equation from any equation not a linear combination of the equations of the simultaneous system. Thus a fundamental condition, Fisher (1966, p. 20), establishing the standard rank condition, which in its turn concerns the inadmissibility of linear combinations of the equations of the system, is not fulfilled.

2.3.1 The general case. For the case where nothing is known about the properties of the process generating the disturbances Fisher (1966) gives a condition that restores the standard rank condition. Rewriting (1) with the exogenous and the lagged endogenous variables separated[1]

$$y_t = \beta y_t + \Gamma_1 y_{t-1} + \Gamma_2 z_t^e + \delta_t \tag{5}$$

the systematic part of y_{t-1} is obtained from the restricted reduced form through repeated substitution as

$$y_{t-1}^* = \sum_{\nu=1}^{\infty} [(I - \beta)^{-1}\Gamma_1]^{\nu-1} (I - \beta)^{-1}\Gamma_2 z_{t-\nu}^e \tag{6}$$

assuming that the dynamic system (5) is stable.

The condition is that the elements of y_{t-1}^* and z_t are not connected by any linear identities which is necessary and sufficient for observational information to distinguish the first equation, say, of (5) from any equation not a linear combination of the equations of (5). In any case, observational information cannot distinguish, unless more conditions are imposed, the true first equation from equations that are linear combinations of the equations of (5).

Mallella and Rao (1971) argue that Fisher's condition is too restrictive and develop alternative assumptions allowing for identities between the elements of y_{t-1}^* and z_t. Their assumptions are: (I) No component of z_t lagged one period appears as component of z_t, (II) there exist no linear identities connecting $z_t, \ldots, z_{t-p}$, p being the number of equations in the model and (III) the model does not imply identities of the type $y_{it}^* = 0$.

Diestler (1975) has recently stated rank-conditions which ensure that the classical identification theorems can be applied. The conditions are given for stationary errors in the general case, as well as in the moving average and autoregressive case.

2.3.2 The first order autoregressive case. Restrictions concerning the form of the process generating the disturbances can be employed studying identifiability criteria. Here we will do this for a special case considering the situation

[1] We are assuming that one period lags only is present.

where the disturbances δ_t are generated according to the first order autoregressive process

$$\delta_t = R\,\delta_{t-1} + \delta_t^+ . \tag{7}$$

The transformed system

$$\begin{aligned} y_t &= \beta y_t + \Gamma z_t + R(I-\beta)y_{t-1} - R\Gamma z_{t-1} + \delta_t^+ = \\ &\beta y_t + \Gamma z_t + \Gamma_1 y_{t-1} + \Gamma_2 z_{t-1} + \delta_t^+ \end{aligned} \tag{8a–b}$$

contains nonlinear restrictions on the parameters β, Γ, Γ_1 and Γ_2. One approach to the problem is then to consider conditions in the nonlinear case and develop the concept of local identifiability, see Fisher (1966, pp. 151–162). This approach is rather complicated and not so well explored.

Another way to consider the case of the first order autoregressive process is to start from the reduced form of (1) or

$$y_t = \Pi z_t + \epsilon_t \tag{9}$$

where the disturbances according to (7) are generated from

$$\epsilon_t = (I-\beta)R(I-\beta)^{-1}\epsilon_{t-1} + \epsilon_t^+ = R_\epsilon \epsilon_{t-1} + \epsilon_t^+ . \tag{10a–b}$$

This and (9) give the transformed reduced form

$$y_t = R_\epsilon y_{t-1} + \Pi z_t - R_\epsilon \Pi\, z_{t-1} + \epsilon_t^+ . \tag{11}$$

Now disregarding for the moment the problem to identify the parameters of (11) the conditions to obtain the SF-parameters are

$$(I-\beta)\,(I\,|\,\Pi) = (I-\beta\,|\,\Gamma) \tag{12}$$

$$(I-\beta)^{-1} R_\epsilon\,(I-\beta) = R\,. \tag{13}$$

If there are no *a priori* restrictions on R and the restrictions on the elements of β and Γ are linear it is realized that the identification problem is identical to that of the case with no serial correlation. The approach is discussed by Sargan (1961) who gives evidence that the resulting model is overidentified when R is assumed diagonal.

The identification of the parameters of (11) remain. In general this problem is similar to the complicated case of nonlinear restrictions above. Sargan (1961) considers some special situations easier to deal with; he gives conditions involving the characteristic roots of R_ϵ in the case where the elements of z_t in (1) are not exogenous.

A final possibility of identification goes through (8b) which, ignoring the restrictions on the parameters, is a linear interdependent system. The device is then to work with this system only and to consider the elements of B, Γ, Γ_1 and Γ_2 as unrestricted. A modification of this approach which is more satisfactory, is to consider the restrictions on the parameters of only one relation at the time. In terms of a limited-information view, cf. limited-information estimation where the specification of the relation under estimation only is considered, this is a reasonable approach suggested by Sargan (1961).

2.4 Testing for serial correlation

In chapter 1 the serious implications of autocorrelated disturbances in regression equations were discussed. It is as a conclusion highly desirable to be able to test for their presence. In this section we will discuss such test for large and small samples. The treatment will not be fully exhaustive as most textbooks on econometrics include discussions on tests for single regression equations, see e.g. Christ (1966), Johnston (1972) and Malinvaud (1970). Their treatment of the simultaneous equation case is however rather summary. Our attention will be directed towards that case and tests related to the estimators that are to be developed.

Suppose, to start with the simpler case, that the estimated version of a single regression equation

$$y_t = \gamma z_t + \epsilon_t \qquad (1)$$

is written

$$y_t = c z_t + e_t \qquad (2)$$

where γ and c are 1 x m-vectors, z_t is m x 1 and t = 1, T. The residuals e_t are estimates of the unobservable disturbances ϵ_t and are available for testing purposes. Let c denote the least squares estimate of the parameter vector γ. It is obvious that the sequence of residuals is dependent not only on the sequence of disturbances but also on the sequence of values taken by the regressor variables. This means that the distribution of the residuals cannot be derived independently of the values of z_t. Provided c is consistent, the residuals converge in probability to the disturbances. This opens the possibility of large sample tests where the residuals are treated as observations on the error variable ϵ_t.

2.4.1 Large sample tests. A first alternative is to apply non-parametric tests to the sequence of residuals. The significance levels of these tests will not be exact for finite samples but given sufficiently large samples they will be sufficiently accurate. An account of various non-parametric tests is e.g. given by Siegel (1956).[1]

A second alternative is to use tests based on theoretical distributions. One example is the ratio of the mean square successive difference to the variance statistic or von Neumann's ratio (1941)

$$\frac{\delta^2}{s^2} = \frac{\Sigma(e_t - e_{t-1})^2/(T-1)}{\Sigma(e_t - \bar{e})^2/T} \tag{3}$$

where $\bar{e}$ is the mean value, which is zero in ordinary regression. The distribution of this test statistic has been tabulated by Hart (1942) on the assumption that the e_t's are random drawings from a normal distribution. Hart's table is not appropriate testing hypotheses concerning the serial correlation of disturbances because it is intended to apply to observable random variables. For large samples the test statistic (3) may be taken as approximately normal with the approximative formulae for the mean and the variance given by Johnston (1972, p. 250):

$$E(\delta^2/s^2) = 2T/(T-1), \quad V(\delta^2/s^2) = 4T^2(T-2)/(T+1)(T-1)^3. \tag{4a–b}$$

These two alternatives utilized the residuals in (2) for testing purposes. A different approach is to test the significance of the parameters of a specified correlation structure of the disturbances. This procedure requires reason to believe that the disturbances are generated e.g. according to

$$\epsilon_t = \rho\, \epsilon_{t-1} + \epsilon_t^+, \; |\rho| < 1 \tag{5}$$

or according to a second order process or a moving average process. Inserting (1) into (5) we get

$$y_t = \rho\, y_{t-1} + \gamma\, z_t - \rho\, \gamma\, z_{t-1} + \epsilon_t^+ \tag{6}$$

which is a regression equation nonlinear in the parameters and with the lagged variable y_{t-1} as regressor. The exact distribution of one of the usual estimators as applied to (6) can be expected to be intractable and one has to be satisfied with asymptotic results. When deriving estimators in chapters 3–5 we consider

[1] The small sample properties of the sign test are discussed by Habibagahi and Pratschke (1972).

a more general situation than (6) and asymptotic tests of the hypothesis that the autocorrelation is zero can be obtained as a special case from that discussion; see also below. Sargan (1964) suggests an asymptotic test for higher order autoregression. This test is based on the fitting of a first order process only. The significance of the parameter ρ can also be examined through a likelihood ratio test. Making appropriate normality assumptions and estimating (1) and (6) by the maximum likelihood method we are able to compare the associated likelihood values L_1 and L_6 to obtain a measure of the contribution by the additional parameter. Asymptotically

$$2(\log L_1 - \log L_6) \sim \chi^2(1) \tag{7}$$

i.e. the difference follows the chi-square distribution with one degree of freedom.

Those were tests for a single relation. Similar tests are possible to derive for simultaneous equations. The classic interdependent system was in section 2.1 specified as[2]

$$y_t = \beta y_t + \Gamma z_t + \delta_t \tag{8}$$

Under the assumption that the disturbances are generated according to a specific structure say

$$\delta_t = R\,\delta_{t-1} + \delta_t^+ \tag{9}$$

we obtain a transformed system

$$y_t = R\,y_{t-1} + \beta\,y_t + \Gamma\,z_t - R\,(\beta y_{t-1} + \Gamma z_{t-1}) + \delta_t^+ \tag{10}$$

For the system (10) estimators will be developed in the following chapters with R as either a diagonal or a full matrix. The asymptotic covariance matrices derived may be used for tests concerning elements or linear combinations of elements of this matrix. For the full-information maximum likelihood estimator we get a likelihood ratio test of the hypothesis R = 0 by comparing the likelihood of (10) L_{10} with that of (8) L_8

$$2\,(\log L_8 - \log L_{10}) \sim \chi^2(k) \tag{11}$$

where k is the difference in dimension between the two specifications, which is p^2 if R is full, cf. Hendry (1971). Sargan (1964) suggests an alternative test that in this case comes out as, cf. Hendry (1971),

[2] Here as in the previous sections it is assumed that the system consists of p behavioral relations only.

$$2(\log L_{RF10} - \log L_{RF8}) \sim \chi^2(ps) \qquad (12)$$

L_{RF8} is the likelihood of the reduced form of (8); unrestricted as regards the observed variables but with the error structure restricted according to (9). L_{RF10} is the likelihood of the unrestricted reduced form of (10) with redundant variables eliminated; s is the number of predetermined non-redundant variables. Redundant variables are those that are exact linear functions of the other predetermined variables. A significant result can be interpreted as implying that the dynamic structure of the model is incorrectly specified.

2.4.2 Small sample tests. Econometrics is in practice usually concerned with short time-series. It is therefore vital to develop exact tests applicable in these situations. The main difficulty already mentioned is that the sequence of residuals depends not only on the disturbances but also on the sequence of values taken by the regressor variables. Durbin and Watson (1950, 1951) succeeded in establishing the lower and upper limits of the significance levels for von Neumann's test. In the terms of (2) they defined the "d"-statistic as

$$\text{"d"} = \frac{\Sigma(e_t - e_{t-1})^2}{\Sigma e_t^2} \qquad (13)$$

and showed for given values of ϵ_t (t = 1, T) that "d" is necessarily contained between two limits d_L and d_U which are independent of the values on the nonstochastic regressor vector z_t. The limits are random variables whose distributions are determined by the number of observations T and the number of regressors m. Durbin and Watson determined the probability densities of d_L and d_U, assuming that the ϵ_t's are homoscedastic and independently and normally distributed, and tabulated values d_1 and d_2 giving the confidence levels of the limits for certain significance levels. Testing the hypothesis of no autocorrelation against the alternative of positive autocorrelation the "d"-value calculated from the residuals is compared to d_1 and d_2. If this "d"-value is smaller than d_1 the hypothesis of independence is rejected. If it is larger than d_2 this hypothesis is not rejected. If the "d"-value obtained is between d_1 and d_2 we are in the inconclusive region and cannot say whether the hypothesis should be rejected or not. Through a simple transformation the alternative hypothesis may be specified as the presence of negative autocorrelation. The "d"-statistic was constructed with the alternative hypothesis of disturbances generated by a stationary Markov process, see (5), in mind. The test is therefore powerful when the disturbances are generated by this process.

One drawback of the Durbin-Watson test is the inconclusive region, the

circumstance that the test does not lead to a conclusion in a certain number of cases. This region can be eliminated or diminished if one is willing to make more precise assumptions concerning the z_t's. Malinvaud (1970, p. 509) contends that the exact value of the significance level of "d" is near d_U if the regressor variables are approximately linear combinations of a constant and of sinusoidal sequences with long periods, more precisely with periods 2T, T, 2T/3,..., 2T/(m-1). He further argues that most exogenous variables have gradual evolutions, i.e. are combinations of sinusoidal sequences with long rather than short periods, and that a value of "d" in the inconclusive region generally indicates significant serial correlation.

Theil and Nagar (1961) have derived limits for the case in which the first and second differences of the explanatory variables are small in absolute value compared with the range of the corresponding variable itself. These assumptions are satisfied by variables with fairly small cyclical components and strong time trends, which are usual in econometrics, but hardly by variables expressed in first difference form. They further show that the upper limit d_U under the assumptions made is approximately equal to the true significance level.

In a recent article Durbin and Watson (1971) examine methods of calculating the exact distribution and approximations to the exact distribution of the "d"-statistic given the actual values on z_t. These involve rather heavy calculations. Durbin and Watson recommend that the exact significance point should be calculated if a "d"-value is contained in the inconclusive region.

A different approach has been the construction of exact tests which probability of rejection do not depend on the values of z_t. This is done by transforming the residuals to a new set of values whose distribution is independent of z_t and calculate appropriate test statistics from these values. Theil has developed the BLUS procedure (1965, 1968, see also 1971), where he suggests an estimator of e_t (t=1,T) which is the Best Linear Unbiased with a Scalar covariance matrix. Durbin (1970B) introduces a second type of exact tests.

Since these tests also involve a substantial amount of calculation it is appropriate to ask how much is gained by the extra effort. Durbin and Watson (1971) show that for alternatives in the neighbourhood of the null hypothesis the test based on the original "d"-statistic is a locally most powerful invariant test. This property is not shared by Theil's and Durbin's (1970) tests and it appears that these compare unfavorably as regards power. Koerts and Abrahamse (1968) have compared the relative powers of the BLUS procedure and the original "d"-statistic through simulation experiments postulating a first-order process (5) for the disturbances. Their results indicate that the BLUS proce-

dure rejects the null hypothesis more often than its competitor but that neither procedure is especially powerful when the number of observations is small and the disturbances are moderately autocorrelated. They have not attempted to deal with the problem of the inconclusive region of the Durbin-Watson test but if "d"-values in that region are taken to indicate serial correlation, the indication of the experiments is reversed with the BLUS procedure as less powerful. It seems that more evidence is needed before the more complicated tests can be claimed superior.

The discussion concerning small sample tests has up till now applied to single equations with fixed regressors. When some of the explanatory variables are lagged dependent variables, tests based on residuals from regression on fixed regressors are invalid even when the null hypothesis of serial independence is true, as shown by Durbin (1970A). He constructs a large sample test which for the model

$$y_t = \beta_1 y_{t-1} + \gamma z_t + \epsilon_t \tag{14a}$$

$$\epsilon_t = \rho \epsilon_{t-1} + \epsilon_t^+ \tag{14b}$$

comes out as testing

$$h = r \sqrt{T/(1-\hat{V}(b_1))} \tag{15}$$

as a standard normal deviate. The difference between this model and (1) is the right-hand side term y_{t-1}. β_1 is a scalar parameter, b_1 the least square estimate obtained from (14a) and r is the serial correlation coefficient obtained regressing the residual e_t on e_{t-1}. $\hat{V}(b_1)$ is the least squares estimate of the variance.

Small sample tests in simultaneous equations in the absence of lagged endogenous variables have been considered by Durbin (1957). Recalling the classic interdependent system

$$y_t = \beta y_t + \Gamma z_t + \delta_t \tag{16}$$

we shall explain Durbin's procedure. He assumes that the model has been estimated by the limited-information method of Anderson and Rubin (1949, 1950) and develops tests for the disturbance of a separate relation. In the case of a just identified relation it turns out that a test may be performed by calculating the Durbin-Watson "d"-statistic and entering the usual tables. Far more frequent are overidentified equations. Durbin arrives at the "d"-statistic applied to the residuals obtained regressing

$$y_{it} - b_{(i)} y_{(i)t} \tag{17}$$

on all the independent variables of the system. $b_{(i)}$ denotes the limited-information estimates of the parameters of the endogenous variables in relation i which is under study and $y_{(i)}$ the corresponding variables. He conjectures that the difference between this test and the usual Durbin-Watson test is small and claims that a good approximative test is obtained that way. When the model contains lagged endogenous variables the tests derived are no longer exact. Durbin (1957, p. 370) claims that they may be expected to hold approximatively but in view of what was said earlier considerable caution is necessary when the i^{th} relation contains y_i lagged as explanatory variable. Nerlove and Wallis (1966), see also Malinvaud (1970, p. 557) show for a simple autoregressive relation that the "d"-statistic is asymptotically biased towards 2. Their results indicate that the Durbin-Watson test is not especially powerful against Markov alternatives when lagged endogenous variables are present. This conclusion is confirmed by sampling experiments conducted by Malinvaud (1961).

2.4.3 Conclusion. It follows from our discussion that the simple Durbin-Watson "d"-statistic is a useful indicator of the serial correlation of the disturbances of a separate relation of an interdependent system, at least provided the specific relation does not contain lagged endogenous variables. As regards the inconclusive region there is evidence, provided the variables behave "regularly", that a calculated "d"-value in the region indicates autocorrelated disturbances.

The large sample relation between "d" and the estimated autocorrelation coefficient r (cf. (5))

$$\text{"d"} = 2(1-r) \tag{18}$$

is the bridge that carries over to some other tests. These are large-sample justified only and presuppose a different testing strategy where the system is estimated with a specified autocorrelation structure and the significance of the parameters of that structure is examined. Following this procedure we have at our disposal tests, not only for the limited-information situation (5), but also for the full-information situation where the vector of disturbances of the system is generated by a vector autoregressive process, see (9). A drawback with the large sample tests is their reliance on a specific correlation structure. It would however be unfair to disqualify the application of these tests on that circumstance without considering the properties of the competitors. The Durbin-Watson test is powerful against the alternative hypothesis of a disturbance term generated by a first order autoregressive process but for other generating processes we expect less power. In some cases the Durbin-Watson test

may be more powerful for other alternatives such as second order Markov processes or moving average processes, as shown by Blattberg (1973), but these alternatives are possible to parametrize and incorporate into the original model. The usual estimators may then be modified to take account of these parameters also and asymptotic covariance matrices and tests may be derived.

Tests with a less restricted category of alternative hypotheses have, it seems, not been considered for simultaneous equations. For single equations Durbin (1969) examines the cumulative periodogram of the residuals which is likely to reveal more complicated correlation structures than those discussed. A lucid grasp of the outcome of the testing procedure is obtained through a graph. This will serve as a remainder of the residual plot which may give a rough idea as to the serial correlation of the disturbances.

CHAPTER 3: THE FULL-INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR

3.1 Introduction

The Full-Information Maximum Likelihood estimator (FIML), of Koopmans, Rubin and Leipnik (1950), is known to possess optimal properties asymptotically. In the standard case, i.e. the case without serially correlated disturbances, it has been proved that the full-information estimators of the identified parameters are consistent, asymptotically normally distributed and asymptotically efficient under general conditions. This in agreement with well-known properties of maximum likelihood methods.

Suggested extensions of the FIML-method considering disturbances generated by an autoregressive process are discussed in subsection 1.5.1 The difference between these earlier approaches and the one to follow lies in the generality. We will give the asymptotic covariance matrix for a general case and within our framework restrictions between all parameters of the model, including those specifying the first order auto-regressive process generating the disturbances, can be handled. A drawback is that extensions to higher order auto-regressive processes are complicated. The treatment results in a computational algorithm, which is a generalization of the treatment of Jöreskog (1973) for the standard case. The algorithm is discussed in section 6.2. In this chapter we will derive the estimator, the associated asymptotic covariance matrix and somewhat discuss the problem of undersized samples.

3.1.1 The model and the treatment of the identities. The basic structural form when we consider the FIML-estimator is written

$$\beta_I y_t - \Gamma z_t = \delta_t \qquad (1)$$

where $\beta_I = I - \beta$ and Γ are parameter matrices of dimension p x p and p x m respectively. y_t is the p x 1 vector of the endogenous and z_t the m x 1 vector of the predetermined variables[1]. The disturbances are written as the p x 1 vector δ_t. Equation (1) applies, as it is specified so far, only to the case of no identities. In the more general case the identities are eliminated according to the following procedure. A system with p behavioral relations

[1] In this chapter we allow for stochastic exogenous variables. The use of β and Γ also for the estimated quantities in the chapter should not cause any confusion.

and q residualfree relations or identities is written by the use of partitioned matrices as

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{Bt} \\ y_{It} \end{pmatrix} - \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} z_t = \begin{pmatrix} \delta_{Bt} \\ 0 \end{pmatrix}, \quad (t=0, T) \qquad (2)$$

where β_{11} is p x p, β_{12} p x q, Γ_1 p x m, β_{21} q x p, β_{22} q x q and Γ_2 q x m, β_{11}, β_{12} and Γ_1 consist of parameters which are unknown whereas the elements of β_{21}, β_{22} and Γ_2 are known constants. We solve for y_{It} in the system of identities

$$\beta_{21} y_{Bt} + \beta_{22} y_{It} - \Gamma_2 z_t = 0 , \qquad (3)$$

provided β_{22} is nonsingular, and insert the solution in the system of behavioral relations in (2). After a slight rearrangement the result is

$$(\beta_{11} - \beta_{12}\beta_{22}^{-1}\beta_{21}) y_{Bt} - (\Gamma_1 - \beta_{12}\beta_{22}^{-1}\Gamma_2) z_t = \delta_{Bt} \qquad (4)$$

which with a notational change can be written as (1). In the sequel of this chapter we shall mean $(\beta_{11} - \beta_{12}\beta_{22}^{-1}\beta_{21})$ by β_I, $(\Gamma_1 - \beta_{12}\beta_{22}^{-1}\Gamma_2)$ by Γ and δ_{Bt} by δ_t. The system (2) and consequently (4) is normalized by the specification that β_{11} and β_{22} have units on their main diagonals.

The disturbances δ_t are assumed to be generated by a stationary first-order autoregressive process

$$\delta_t = R\,\delta_{t-1} + \delta_t^+ , \quad (t=1, T) . \qquad (5)$$

Inserting (1) we obtain the transformed system

$$\beta_I y_t - \Gamma z_t - R(\beta_I y_{t-1} - \Gamma z_{t-1}) = \delta_t^+ , \quad (t=1, T) \qquad (6)$$

after some modification. We define the matrices $A = (\beta_I \vdots - \Gamma)$, $x_t' = (y_t' \vdots z_t')$, $X = (x_1, \ldots, x_T)$, $\Delta^+ = (\delta_1^+, \ldots, \delta_T^+)$ and write (6) for all timepoints as

$$A X - R A X_l = \Delta^+ \qquad (7)$$

where X_l is X lagged.

3.2 The FIMLA estimator

We assume that β_I is nonsingular and define the Reduced Form (RF) by multiplying (3.1.6) by β_I^{-1} to the left. The RF-disturbances $\epsilon_t^+ = \beta_I^{-1}\,\delta_t^+$ are assu-

med to be jointly normally distributed with mean vector zero and a nonsingular covariance matrix Σ. The covariance matrix of the SF-disturbances is denoted Ω ; $\Sigma = B_I^{-1} \Omega B_I^{-1\prime}$. When the ϵ_t^+'s (t=1, T) are independently distributed their likelihood is written

$$L(\epsilon_1^+, ..., \epsilon_T^+) = (2\pi)^{pT/2} (\det\Sigma)^{-T/2} \exp\{ -1/2 \sum_t \epsilon_t^{+\prime} \Sigma^{-1} \epsilon_t^+ \} \qquad (1)$$

Through a linear transformation we obtain the likelihood of the dependent variables as[1]

$$L(y_1, ..., y_T | z_1, ..., z_T) = (2\pi)^{-pT/2} (\det B_I^{-1} \Omega B_I^{-1\prime})^{-T/2}$$
$$\exp(-1/2 \sum_t (B_I^{-1}(Ax_t - RAx_{t-1}))' B_I' \Omega^{-1} B_I (B_I^{-1}(Ax_t - RAx_{t-1}))). \qquad (2)$$

The log likelihood is written

$$\log L = - pT/2 \log(2\pi) + T/2 \log \det\Omega^{-1} + T \log \det B_I - 1/2 \operatorname{tr}(AX - - RAX_l)' \Omega^{-1} (AX - RAX_l) \qquad (3)$$

The *Full-Information Maximum Likelihood estimator in the case of Autoregressive disturbances (FIMLA)* is defined by maximizing logL with respect to the unknown elements of Ω, A and R.

The partial derivatives of logL are obtained using matrix differentials, for the technique cf. Neudecker (1969, p. 961). Our interest is first concentrated to Ω which can be eliminated from (3). One simple way to show this is to regard the other parameters, except those in Ω, as constants for a moment. Then differentiation of logL gives

$$d\log L = T/2 \operatorname{tr} \Omega d\Omega^{-1} - 1/2 \operatorname{tr}(AX - RAX_l)(AX - RAX_l)' d\Omega^{-1} \qquad (4)$$

and the likelihood function

$$\Omega = 1/T (AX - RAX_l)(AX - RAX_l)' = 1/T \Delta^+ \Delta^{+\prime} \qquad (5)$$

follows. In the derivation of (4) the result tr GH = tr HG was used. This together with the property $d\Omega^{-1} = -\Omega^{-1} d\Omega \Omega^{-1}$ gives (5).

Inserting the FIMLA estimator of Ω as defined in (5) in (3) the concentrated likelihood

$$\log L = - pT/2 \log(2\pi) - T/2 \log \det\Omega + T \log \det B_I \qquad (6)$$

is obtained. It is possible to concentrate this expression further eliminating R

[1] The Jacobian $J = (\delta \epsilon_{it}^+ / \delta y_{jt})$ is identical to the identity matrix since $\epsilon_t^+ = y_t - B_I^{-1}(\Gamma z_t - R(B_I y_{t-1} - \Gamma z_{t-1}))$. Thus the determinant of J need not be explicitly written in (2).

in a process similar to that leading to (4). The result appealing from mathematical considerations but not from computational, since restrictions in R then are difficult to handle, will not be used here. We only write the estimator obtained

$$R = AXX_l'A'(AX_lX_l'A')^{-1} \tag{7}$$

for the sake of completeness. The perceptive reader may have observed that the i^{th} row of R is obtained as the regression of the i^{th} disturbance term δ_{it} on δ_{it-1}, (i=1, p).

Maximizing (6) is equivalent to minimizing F where F is defined from

$$2F = \log \det \Omega - \log(\det B_I)^2 \tag{8}$$

Ω is given in (5) which is rewritten as

$$\Omega = AM_{00}A' - AM_{10}'A'R' - RAM_{10}A' + RAM_{11}A'R' \tag{9}$$

where M_{00}, M_{10} and M_{11} are matrices of moments e.g. $M_{10} = 1/T\ X_lX'$.

Differentiation of (8) yields

$$2\,dF = tr(\Omega^{-1}\,d\Omega - 2B_I^{-1}dB_I) \tag{10}$$

Remembering that

$$A = (B_{11} - B_{12}B_{22}^{-1}B_{21}' - \Gamma_1 + B_{12}B_{22}^{-1}\Gamma_2) \tag{11}$$

which can be written

$$A = A_1 - B_{12}B_{22}^{-1}A_2 \tag{12}$$

where $A_1 = (B_{11}' - \Gamma_1)$ and $A_2 = (B_{21}' - \Gamma_2)$ we obtain $d\Omega$, dB_I and finally dF expressed in dA_1, dB_{12} and dR as

$$\begin{aligned}
2dF = {} & tr[(AM_{00} - RAM_{10})'\,\Omega^{-1} - (AM_{10}' - RAM_{11})'\,\Omega^{-1}R]\,dA_1 + \\
& + tr[\Omega^{-1}(AM_{00} - RAM_{10}) - R'\Omega^{-1}(AM_{10}' - RAM_{11})]\,dA_1' - \\
& - tr[B_{22}^{-1}\,A_2[(AM_{00} - RAM_{10})'\Omega^{-1} - (AM_{10}' - RAM_{11})'\Omega^{-1}R]]\,dB_{12} - \\
& - tr[[\Omega^{-1}(AM_{00} - RAM_{10}) - R'\Omega^{-1}(AM_{10}' - RAM_{11})]\,A_2'\,B_{22}^{-1'}]\,dB_{12}' - \\
& - tr[(AM_{10}'A' - RAM_{11}A')'\,\Omega^{-1}]\,dR - \\
& - tr[\Omega^{-1}\,(AM_{10}'A' - RAM_{11}A')]\,dR' - \\
& -2tr\,B_I^{-1'}\,dB_{11}' + 2tr\,B_I^{-1'}\,B_{21}'\,B_{22}'^{-1}\,dB_{12}'
\end{aligned} \tag{13}$$

Utilizing that dtr GA_1 = tr G dA_1 implies dtr G $A_1/dA_1 = G'$ where G is a matrix of suitable order the first order partial derivatives of F are obtained as

$$\frac{dF}{dA_1} = \Omega^{-1}(AM_{00} - RAM_{10}) - R'\Omega^{-1}(AM'_{10} - RAM_{11}) - (\beta'^{-1}_I \vdots 0) \tag{14}$$

$$\frac{dF}{d\beta_{12}} = -\Omega^{-1}(AM_{00} - RAM_{10})A'_2\beta'^{-1}_{22} + R'\Omega^{-1}(AM'_{10} - \\ - RAM_{11})A'_2\beta'^{-1}_{22} + \beta'^{-1}_I\beta'_{21}\beta'^{-1}_{22} \tag{15}$$

$$\frac{dF}{dR} = -\Omega^{-1}AM'_{10}A' + \Omega^{-1}RAM_{11}A' \tag{16}$$

Equated to zero (14)–(16) define the maximum likelihood estimator of the elements of A_{11}, β_{12} and R. In the case of several local maxima we discriminate between these through the associated likelihood values.

3.3 Restrictions and the asymptotic distribution

Linear and nonlinear restrictions among the parameters can be handled as follows, cf. e.g. Jöreskog (1973). Let the s x 1 vector θ denote the unrestricted parameters of the model in (3.1.7). Write all non-zero elements of A_1, β_{12}, and R as the u x 1 vector μ. Each of these elements may be a known function of the parameters θ to be estimated

$$\mu_i = f_i(\theta), (i=1, u) . \tag{1}$$

Then F as defined in (3.2.8) is a function of the basic parameters in θ, denoted $H(\theta)$. The first order derivatives of H

$$\frac{dH}{d\theta} = K'\frac{dF}{d\mu} \tag{2}$$

are utilized to define the estimator. K is the u x s matrix whose ij^{th} element is $df_i/d\theta_j$.

The asymptotic distribution of the FIMLA estimator of the unrestricted parameter vector is obtained according to the following theorem.

Theorem 1. Under quite general conditions–that include the assumptions of section 2.1, the assumption of normally distributed disturbances for the transformed system and certain regularity assumptions–the limiting distribution of $T^{1/2}(\hat{\theta} - \theta)$, where $\hat{\theta}$ denotes the maximum likelihood estimator of the un-

restricted parameter vector and θ the corresponding population values, is normal

$$T^{1/2}(\hat{\theta} - \theta) \sim N(0, \text{plim } T(\frac{d^2 \log L}{d\theta d\theta'})^{-1}) \tag{3}$$

with the associated mean vector and covariance matrix given within the parentheses.

We will not prove this theorem. The reader is referred to Cramér (1946, pp. 500–503), Koopmans, Rubin and Leipnik (1950, p. 133 et seq.) and Christ (1966, pp. 427–428) where the asymptotic distributions of maximum likelihood estimators are discussed in more detail. Log L as used in theorem 1 is defined in (3.2.6). In terms of the original restricted parameters as contained in μ, the matrix of the second order derivatives is

$$\frac{d^2 H}{d\theta d\theta'} = K' \frac{d^2 F}{d\mu d\mu'} K \tag{4}$$

3.3.1 An explicit expression of the asymptotic covariance matrix. It is a rather laborious task to derive the matrix of the second order partial derivatives of F and from that an explicit expression of the asymptotic covariance matrix. The point of departure is the matrices

$$\frac{dF}{dA_1} = U \quad \frac{dF}{d\beta_{12}} = V \quad \frac{dF}{dR} = W \tag{5}$$

which we differentiate. The first matrix is obtained as

$$\begin{aligned} dU = {} & d(\Omega^{-1})(AM_{00} - RAM_{10}) + \Omega^{-1}(d(A)M_{00} - d(R)AM_{10} - Rd(A)M_{10}) - \\ & - (d(R')\Omega^{-1} + R'd(\Omega^{-1}))(AM'_{10} - RAM_{11}) - R'\Omega^{-1}(d(A)M'_{10} - \\ & - d(R)AM_{11} - Rd(A)M_{11}) - (d\beta_I^{-1'} \nmid 0). \end{aligned} \tag{6}$$

By developing $d\Omega^{-1}$ from (3.2.9) and inserting the result, we will obtain the second order derivatives where elements in A_1 are involved.

In the following the short notations

$$\begin{aligned} G_1 &= AM_{00} - RAM_{10} \\ G_2 &= AM'_{10} - RAM_{11} \end{aligned} \tag{7}$$

are used. The derivatives are identified according to the rule where

$$\begin{aligned} dF/dS &= Q \\ dQ &= H_1 dS H_2 + H_3 dS' H_4 \end{aligned} \tag{8}$$

imply that $$\frac{d^2F}{ds_{gh}ds_{ij}} = (H_1)_{gi}(H_2)_{jh} + (H_3)_{gj}(H_4)_{ih} \cdot \qquad (9)$$

F is here a function of the elements of the matrix S and H_1 to H_4 are matrices of constants. $(H_1)_{gi}$ denotes the gi^{th} element of H_1. Writing the gh^{th} element of A_1 as a^1_{gh} we have

$$\frac{d^2F}{da^1_{gh}da^1_{ij}} = -\Omega^{gi}(G_1'\Omega^{-1}G_1 - M_{00})_{jh} + (R'\Omega^{-1})_{gi}(G_1'\Omega^{-1}G_2 - M'_{10})_{jh} +$$
$$+ (\Omega^{-1}R)_{gi}(G_2'\Omega^{-1}G_1 - M_{10})_{jh} - (R'\Omega^{-1}R)_{gi}(G_2'\Omega^{-1}G_2 - M_{11})_{jh} -$$
$$- (\Omega^{-1}G_1)_{gj}(\Omega^{-1}G_1)_{ih} + (\Omega^{-1}G_2)_{gj}(R'\Omega^{-1}G_1)_{ih} + (R'\Omega^{-1}G_1)_{gj}$$
$$(\Omega^{-1}G_2)_{ih} - (R'\Omega^{-1}G_2)_{gj}(R'\Omega^{-1}G_2)_{ih} + \kappa(\beta_I^{jg}\beta_I^{hi}) \qquad (10)$$

where Ω_{gi} denotes the gi^{th} element of Ω and Ω^{gi} the gi^{th} of Ω^{-1}. κ is a scalar which is 0 when gh or ij denote a Γ-element, otherwise κ is 1.

The other second order derivatives where A_1 is involved come out as

$$\frac{d^2F}{da^1_{gh}d\beta^{12}_{ij}} = \Omega^{gi}(\beta_{22}^{-1}A_2(G_1'\Omega^{-1}G_1 - M_{00}))_{jh}$$
$$- (R'\Omega^{-1})_{gi}(\beta_{22}^{-1}A_2(G_1'\Omega^{-1}G_2 - M'_{10}))_{jh}$$
$$- (\Omega^{-1}R)_{gi}(\beta_{22}^{-1}A_2(G_2'\Omega^{-1}G_1 - M_{10}))_{jh}$$
$$+ (R'\Omega^{-1}R)_{gi}(\beta_{22}^{-1}A_2(G_2'\Omega^{-1}G_2 - M_{11}))_{ih}$$
$$+ (\Omega^{-1}G_1A_2'\beta_{22}^{-1\prime})_{gj}(\Omega^{-1}G_1)_{ih} - (\Omega^{-1}G_2A_2'\beta_{22}^{-1\prime})_{gj}$$
$$(R'\Omega^{-1}G_1)_{ih} - (R'\Omega^{-1}G_1A_2'\beta_{22}^{-1\prime})_{gj}(\Omega^{-1}G_2)_{ih}$$
$$+ (R'\Omega^{-1}G_2A_2'\beta_{22}^{-1\prime})_{gj}(R'\Omega^{-1}G_2)_{ih} - \kappa(\beta_{22}^{-1}\beta_{21}\beta_I^{-1})_{jg}\beta_I^{hi} \qquad (11)$$

where β^{12}_{ij} is the ij^{th} element of β_{12}, and

$$\frac{d^2F}{da^1_{gh}d\rho_{ij}} = \Omega^{gi}(AG_2'\Omega^{-1}G_1 - AM_{10})_{jh}$$
$$- (R'\Omega^{-1})_{gi}(AG_2'\Omega^{-1}G_2 - AM_{11})_{jh}$$
$$+ (\Omega^{-1}G_2A')_{gj}(\Omega^{-1}G_1)_{ih}$$
$$- (R'\Omega^{-1}G_2A')_{gj}(\Omega^{-1}G_2)_{ih} - I_{gj}(\Omega^{-1}G_2)_{ih} \qquad (12)$$

where ρ_{ij} is the ij^{th} element of R and I is the identity matrix.

Remaining now are the second order derivatives' matrices which we derive through differentiation of V and W as defined in (5). The result is

$$\frac{d^2F}{d\beta^{12}_{gh}d\beta^{12}_{ij}} = -\Omega^{gi}(B^{-1}_{22}A_2(G'_1\Omega^{-1}G_1 - M_{00})A'_2B^{-1'}_{22})_{jh} +$$
$$+(R'\Omega^{-1})_{gi}(B^{-1}_{22}A_2(G'_1\Omega^{-1}G_2 - M'_{10})A'_2B^{-1'}_{22})_{jh} +$$
$$+(\Omega^{-1}R)_{gi}(B^{-1}_{22}A_2(G'_2\Omega^{-1}G_1 - M_{10})A'_2B^{-1'}_{22})_{jh} -$$
$$-(R'\Omega^{-1}R)_{gi}(B^{-1}_{22}A_2(G'_2\Omega^{-1}G_2 - M_{11})A'_2B^{-1'}_{22})_{jh} -$$
$$-(\Omega^{-1}G_1A'_2B^{-1'}_{22})_{gj}(\Omega^{-1}G_1A'_2B^{-1'}_{22})_{ih} +$$
$$+(\Omega^{-1}G_2A'_2B^{-1'}_{22})_{gj}(R'\Omega^{-1}G_1A'_2B^{-1'}_{22})_{ih} +$$
$$+(R'\Omega^{-1}G_1A'_2B^{-1'}_{22})_{gj}(\Omega^{-1}G_2A'_2B^{-1'}_{22})_{ih} -$$
$$-(R'\Omega^{-1}G_2A'_2B^{-1'}_{22})_{gj}(R'\Omega^{-1}G_2A'_2B^{-1'}_{22})_{ih} +$$
$$+(B^{-1}_{22}B_{21}B^{-1}_{I})_{jg}(B^{-1}_{22}B_{21}B^{-1}_{I})_{hi} \qquad (13)$$

$$\frac{d^2F}{d\beta^{12}_{gh}d\rho_{ij}} = -\Omega^{gi}((AG'_2\Omega^{-1}G_1 - AM_{10})A'_2B^{-1'}_{22})_{jh} +$$
$$+(R'\Omega^{-1})_{gi}((AG'_2\Omega^{-1}G_2 - AM_{11})A'_2B^{-1'}_{22})_{jh} -$$
$$-(\Omega^{-1}G_2A')_{gj}(\Omega^{-1}G_1A'_2B^{-1'}_{22})_{ih} +$$
$$+(R'\Omega^{-1}G_2A')_{gj}(\Omega^{-1}G_2A'_2B^{-1'}_{22})_{ih} +$$
$$+I_{gj}(\Omega^{-1}G_2A'_2B^{-1'}_{22})_{ih} \qquad (14)$$

$$\frac{d^2F}{d\rho_{gh}d\rho_{ij}} = -\Omega^{gi}(AG'_2\Omega^{-1}G_2A' - AM_{11}A')_{jh} -$$
$$-(\Omega^{-1}G_2A')_{gj}(\Omega^{-1}G_2A')_{ih} \qquad (15)$$

Deriving the probability limits of the second order derivatives to obtain the inverse of the information matrix we make use of the following result (see e.g. Wilks 1962, p. 103). If $G(\mu_t)$ is a continous function of the vector of random variables $\mu'_t = (\mu_{1t}, \mu_{2t}, \dots, \mu_{mt})$ which converges in probability to the finite limit, plim $\mu_t = \mu$, then plim $G(\mu_t) = G(\mu)$. We also use that plim$(\alpha\beta)$ = plim(α) plim(β), which holds if α and β have finite probability limits.

The following notations are introduced: plim $T^{-1}ZZ' = \Phi$, plim $T^{-1}Z_lZ' = H$, plim $T^{-1}\Delta_lZ' = K$, plim $T^{-1}\Delta\Delta' = \Psi$, $\beta_I^{-1}\Gamma = \Pi$, $\beta_I^{-1}\Psi\beta_I^{-1'} = \Psi^*$, plim $T^{-1}Z_l\Delta' = L$ and plim $T^{-1}\Delta^+\Delta^{+\prime} = \Omega$.

Z is the m x T matrix of the predetermined variables and Δ and Δ^+ are p x T matrices of disturbances. The subscript l means that the variables in the corresponding matrix are lagged one period. It is now straightforward to derive the probability limits of the moment matrices M_{00}, M_{10} and M_{11}. We get

$$\left.\begin{matrix}\text{plim } M_{11}\\ \text{plim } M_{00}\end{matrix}\right\} = \begin{pmatrix} \Pi\Phi\Pi' + \Pi K'R'\beta_I^{-1'} + \beta_I^{-1}RK\Pi' + \Psi^* & \Pi\Phi + \beta_I^{-1}RK \\ \Phi\Pi' + K'R'\beta_I^{-1'} & \Phi \end{pmatrix}$$

$$\text{plim } M_{10} = \begin{pmatrix} \Pi H\Pi' + \Pi L\beta_I^{-1'} + \beta_I^{-1}K\Pi' + \beta_I^{-1}\Psi R'\beta_I^{-1'} & \Pi H + \beta_I^{-1}K \\ H\Pi' + L\beta_I^{-1'} & H \end{pmatrix} \tag{16, 17}$$

The probability limits of the functions G_1 and G_2 defined in (7) come out as

$$\begin{aligned} \text{plim } G_1 &= (\Omega\beta_I^{-1'} \vdots 0) \\ \text{plim } G_2 &= (0 \vdots 0) \end{aligned} \tag{18}$$

Making use of the notations

$$\begin{aligned} (A_{gi}\beta_{Ijh}) &= A \otimes \beta_I' \\ (\Pi' \vdots I_m) &= P' \end{aligned} \tag{19}$$

where $\otimes$ is the Kronecker product and I_m the identity matrix, we get

$$\text{plim } \frac{d^2F}{da_{gh}^1 da_{ij}^1} = \Omega^{-1} \otimes (P\Phi P' + \begin{pmatrix} \beta_I^{-1}R\Psi R'\beta_I^{-1'} + \Pi K'R'\beta_I^{-1'} + \beta_I^{-1}RK\Pi' & \beta_I^{-1}RK \\ K'R'\beta_I^{-1'} & 0 \end{pmatrix})$$

$$- (R'\Omega^{-1}) \otimes (PHP' + \begin{pmatrix} \Pi L\beta_I^{-1'} + \beta_I^{-1}\Psi R'\beta_I^{-1'} & \beta_I^{-1}K \\ L\beta_I^{-1'} & 0 \end{pmatrix})$$

$$- (\Omega^{-1}R) \otimes (PH'P' + \begin{pmatrix} \beta_I^{-1'}L'\Pi' + \Pi K'\beta_I^{-1} + \beta_I^{-1}R\Psi\beta_I^{-1'} & \beta_I^{-1}L' \\ K'\beta_I^{-1'} & 0 \end{pmatrix})$$

$$+ (R'\Omega^{-1}R) \otimes (P\Phi P' + \begin{pmatrix} \Pi K'R'\beta_I^{-1'} + \beta_I^{-1'}RK\Pi + \Psi^* & \beta_I^{-1}RK \\ K'R'\beta_I^{-1'} & 0 \end{pmatrix})$$

(20)

As is seen comparing (10) and (11) we get plim $\frac{d^2F}{da^1_{gh}\, d\beta^{12}_{ij}}$ multiplying each matrix to the right in the Kronecker products in (20) by $-A_2'B_{22}^{-1'}$ to the right. Multiplying the so obtained matrices by $-B_{22}^{-1}A_2$ to the left yields plim $\frac{d^2F}{d\beta^{12}_{gh}d\beta^{12}_{ij}}$ as is seen from (13). Lacking are now the terms where derivatives with respect to elements of R are involved. We obtain

$$\text{plim}\ \frac{d^2F}{da^1_{gh}d\rho_{ij}} = -\Omega^{-1} \otimes \begin{pmatrix} \Pi K' + & \beta_I^{-1}R\Psi \\ K' & \end{pmatrix} + R'\Omega^{-1} \otimes \begin{pmatrix} \Pi K'R' + \beta_I^{-1}\Psi \\ K'R' \end{pmatrix}$$

(21)

$$\text{plim}\ \frac{d^2F}{d\rho_{gh}d\rho_{ij}} = \Omega^{-1} \otimes \Psi \qquad (22)$$

The final probability limit plim $\frac{d^2F}{d\beta^{12}_{gh}d\rho_{ij}}$ is obtained multiplying the matrices in (21) to the right in the Kronecker products by $-B_{22}^{-1}A_2$ to the left as is seen from (14).

Comment 1: An immediate corollary of (21) is that R = 0 generally leads to asymptotic loss of efficiency if this fact is not used in the estimation procedure.

3.4 Some further points

The desirable asymptotic properties of the full-information method are no guarantee for a good performance in finite samples where other methods may work better. This and the problems of the application of the FIMLA-method in small samples are reasons for considering other estimators. Some alternatives will be discussed in the following chapters. Finite sample properties are discussed in chapter 7.

There is a problem of degrees of freedom for the estimation by full-information maximum likelihood. Klein (1969) suggests conditions concerning sample

size in the standard case. The most important conditions are that the number of observations on the endogenous variables should be larger than the number of unknown parameters ($pT > \Sigma(n_i + m_i)$) and that the number of variables should be smaller than the number of observation points ($T > p + m$). Both conditions apply to the system obtained after elimination of the identities. When the first condition not is fulfilled the estimated covariance matrix of the disturbances is singular and a maximum of the likelihood function will not exist. The implication of the second condition is similar: the covariance matrix of the disturbances $\Psi = 1/T\ AXX'A'$ (cf. 3.2.5) is not nonsingular unless the matrix of moments XX' is positive definite.[1]

In our case of autocorrelation we may expect that the problems of sample size become more accentuated. The unknown parameters are not only the $\Sigma(n_i + m_i)$ structural parameters but also the unknown elements of R which are p^2 in number when this matrix is full. The consequence is that the first condition is sharpened. As regards the second condition we develop the covariance matrix in (3.2.5) as

$$\Omega = 1/T\ (AX - RAX_l)\,(AX - RAX_l)' = 1/T\ CWW'C' \qquad (1)$$

where $C = (A \mid -R\,A)$ and $W = (X' \mid X_l')'$. The order of the moment matrix WW' is $2p + 2m$. In the case of lagged endogenous variables in X this moment matrix is singular but this does not in general mean that Ω is singular since the elements of C are restricted. These restrictions make it difficult to give rank conditions concerning WW' that assure Ω's nonsingularity.

[1] Conditions concerning the restrictions on A are implicitly ignored by Klein.

CHAPTER 4: STRUCTURAL AND REDUCED FORM ORIENTED ESTIMATORS

4.1 Introduction

In this and the following chapter two classes of estimators, both based on re-specified interdependent systems, are developed. The two and three stage least squares methods are obtained as members of the first class, while the GEID-estimator is obtained as member of the second. Rewriting the REID-system (2.1.9)

$$y_t = \beta\eta_t^* + \Gamma z_t + \epsilon_t \tag{1}$$

we have the starting point for this chapter, where the disturbances are assumed non-autocorrelated. The i^{th} behavioral relation of (1) is written for all observations simultaneously as

$$Y_i = (H^*_{(i)}\ Z_{(i)})(\beta_{(i)}\ \gamma_{(i)})' + \epsilon_i \tag{2}$$

where Y_i and ϵ_i are T x 1 vectors, $H^*_{(i)}$ and $Z_{(i)}$ of order T x n_i and T x m_i respectively are matrices containing the "explanatory" variables, and $\beta_{(i)}$ and $\gamma_{(i)}$ of order 1 x n_i and 1 x m_i respectively as before denote the unknown structural parameters. The relation (2) is compactly written as

$$Y_i = \chi_i a_i + \epsilon_i \tag{3}$$

If $H^*_{(i)}$ and thus χ_i were known we could use the ordinary least squares (OLS) estimator a_i defined from

$$\chi_i' Y_i = \chi_i' \chi_i\ a_i \tag{4}$$

to, provided $E(\epsilon_i\epsilon_i') = \xi_{ii}\ I_T$ where I_T is the identity matrix of order T and ξ_{ii} a scalar, obtain efficient estimates of the parameters under general conditions.

Now the main difficulty is that $H^*_{(i)}$ is unobservable. According to subsection 2.1.4 the unrestricted reduced form (URF) and the restricted reduced form (RRF) coincide for the REID specification i.e.

$$\eta_t^* = \Pi z_t = (I - \beta)^{-1}\Gamma z_t\ ,\ (t=1,\ T) \tag{5a–b}$$

where all the parameters are population quantities. The estimated reduced forms however, the projection of Y on Z on the one hand, and the reduced form as obtained from the structural parameters on the other, may differ.

This opens the possibility of two different approaches defining estimators for the system (1); one based on the RRF and another based on the URF.

We call the approach based on the restricted reduced form *structural form oriented* and write

$$X_i^S = (Y_{(i)}^* Z_{(i)}) \tag{6a}$$

for the T x $(n_i + m_i)$ data matrix obtained accordingly. The second approach will be called *reduced form oriented* and we write

$$X_i^R = (Y_{(i)}^* Z_{(i)}) \tag{6b}$$

for the T x $(n_i + m_i)$ matrix where $Y_{(i)}^*$ is obtained from the unrestricted reduced form. An important advantage of the structural form oriented estimators is that they do not require that the number of observations is at least as large as the number of predetermined variables. The methods can be applied when the reduced form oriented methods break down in an unrestricted reduced form step.

4.2 Two limited-information methods

Ignoring the correlations between errors of different relations we will define two limited-information methods. The reduced form oriented method obtained is identical to the Two Stage Least Squares (2SLS) method of Basmann (1957) and Theil (1958). Using on the other hand the restricted reduced form, we obtain the GEneral InterDependent systems (GEID) estimator, Wold (1965) and Ågren (1972). This method was under the name the Fix-Point method considered by Wold as an iterative process. Ågren defines the estimator as the solutions to a set of nonlinear equations. With his approach the properties of the estimator are easier to derive and the difficulties of a process that does not always converge or that may converge to different final values depending on the initial values, are theoretically eliminated. We will use Ågrens definition. The estimated parameters of a single relation i are denoted a_i below–cf. (3) and (4)–the $\Sigma(n_i + m_i)$ row vector a' is $(a'_1,..., a'_p)$.

Definition 1. A structural form oriented estimator, *the GEID estimator,* is defined as the real solutions a_i (i=1,p), if any, of the system

$$X_i^{S\prime} Y_i = X_i^{S\prime} X_i^S a_i, \quad (i=1,p) \tag{1a}$$

$$Y^{*\prime} = (I - B)^{-1} C Z' \tag{1b}$$

where X_i^S is defined in (4.1.6a) and Z is the T x m matrix of observed values of the predetermined variables.

Eliminating Y^* in (1a) using (1b) we get a system which is nonlinear in the unknown parameters. This opens the possibility of several real solutions. We therefore give the following definition, cf. Ågren (1972) and (1973).

Definition 2. Any real solution of (1a–b) will be called an estimator in the *loose sense.* The real solutions belonging to a sequence of solutions that converge in probability to the true parameter values will be called estimators in the *strict sense* if any such exists.[1]

Definition 3. A reduced form oriented estimator, *the 2SLS estimator,* is defined as the solution a_i (i=1,p), if any, of the system

$$X_i^{R\prime} Y_i = X_i^{R\prime} X_i^R a_i \ , \qquad (i=1,p) \qquad (2a)$$

$$Y_i^* = Z(Z'Z)^{-1}Z'Y_i \ , \qquad (i=1,n) \qquad (2b)$$

where Z is a T x m matrix of observed values of the predetermined variables and X_i^R is defined in (4.1.6b).

Comment 1. It is obvious that the least squares steps (1a) and (2a) should be applied to the p behavioral relations only. The Y_i^* -values, on the other hand, are obtained for all n endogenous variables through (1b) and (2b).

Comment 2. The number of observations T required for estimation is different for the two estimators. The reduced form oriented estimator requires that T is at least as large as m , the number of predetermined variables, whereas this is not required for the structural form oriented estimator.

Comment 3. The one to one correspondence between REID and CLID systems, cf. section 2.1.2, implies that the estimators although defined for REID specified systems share the same properties for CLID systems.

Comment 4. The interpretation of 2SLS is here the historical and established view of the method as a double regression method where the first step consists in purging from the explanatory current endogenous variables stochastic components.

[1] The problem of discriminating between solutions is discussed in section 6.4.

We will now discuss the properties of the estimators. The existence of the 2SLS estimator is implied by the nonsingularity of $Z'Z$ and $X_i^{R\prime} X_i^R$, (i=1, p). Consistency and asymptotic normality can be established in a similar way as they are established in a more general case in the following section. We are here content to summarize the results in a theorem.

Theorem 1. Under the assumptions 1–10, where however the disturbances need not be identically distributed as in assumption 9, the 2SLS method is a consistent estimator of the i^{th} relation of a REID-specified system provided plim $T(X_i^{R\prime} X_i^R)^{-1}$ exists as a nonstochastic matrix with finite elements. Adding assumption 11 and assuming that the disturbances are identically distributed the asymptotic distribution is normal

$$T^{1/2}(a_i - a_i) \sim N(0, \psi_{ii}\text{plim}T(\chi_i'\chi_i)^{-1}), \quad (i=1,p) \tag{3}$$

The probability limits of $T(X_i^{R\prime} X_i^R)^{-1}$ and $T(\chi_i' \chi_i)^{-1}$ are identical and $E(\delta_i \delta_i') = \psi_{ii} I_T$.

The asymptotic distribution of the 2SLS estimator for a certain relation, or the joint distribution for all relations, is given in many textbooks on econometric methods.

Lyttkens, see e.g. Mosbaek-Wold (1970), Ågren (1972, 1974) and Edgerton (1973) have under somewhat different sets of assumptions derived the asymptotic properties of the GEID estimator. We will follow the approach of the two latter stating conditions for consistency.

It was mentioned earlier that a system nonlinear in the unknown parameters is obtained eliminating Y^* in (1a) using (1b). This system which defines the GEID estimator is compactly written as

$$F(a, M) = 0 \tag{4}$$

where F is a vector of n_p functions, n_p being the number of unknown parameters in B and Γ. The n_p x 1 vector a contains the corresponding elements in B and C. M is a column vector containing moments $T^{-1}Y_i' Z_j$ and $T^{-1}Z_i'Z_j$. The only functions of Y and Z that appear in (4) are second order sample moments. Under the assumptions given these converge to the corresponding population moments in probability. The use of the implicit function theorem will then, under certain regularity conditions,[2] lead to the existence of a sequence of GEID estimates in the strict sense. This holds for the assumptions

[2] An essential regularity condition is that F should be a continous function of a and M and possess continous first order partial derivatives.

given establishing consistency of the 2SLS method but those assumptions can be weakened. The GEID-estimator thus is consistent under the GEID-specification of section 2.1.4, which the 2SLS estimator in general is not. The GEID estimator is the only estimator considered in this monograph that consistently estimates GEID-specified systems.

The remaining task in this section is to give the asymptotic distribution of the GEID estimator. Now when consistency is established the distribution is derived in a similar way as is the limiting distribution of the 2SLS estimator. This will be evident from the following sections where the asymptotic distribution is derived in the full-information case. We are here content to give the theorem without proof.

Theorem 2. Under the assumptions 1–11 the asymptotic distribution of the GEID estimator is obtained as

$$T^{1/2}(a - \alpha) \sim N(0, \operatorname{plim} T[\chi'((I - \beta)^{-1} \otimes I_T]\chi)^{-1} \operatorname{plim} T^{-1}\chi'(\Xi \otimes I_T)\chi$$
$$\operatorname{plim} T[\chi'((I - \beta)^{-1} \otimes I_T)\chi]^{-1'}) \qquad (5)$$

provided the probability limits in (5) exist as nonstochastic matrices with finite elements. It is also required that F in (4) obey certain regularity conditions. New notations are $\chi = \operatorname{diag}(\chi_1, \ldots, \chi_p)$ and $E(\epsilon\epsilon') = \Xi \otimes I_T$ where $\epsilon' = (\epsilon_1', \ldots, \epsilon_p')$ and $E(\epsilon_{it}\epsilon_{jt}) = \xi_{ij}$.

Comment 5. Lyttkens (1973) has demonstrated that neither the 2SLS nor the GEID estimator is asymptotically more efficient in general. Which is more efficient in a specific situation depends on the structure.

Comment 6. Although interpreted as a limited-information method the GEID estimator has full-information features; see (1b) where Y^* is defined in the structural parameters. This is reflected in the asymptotic covariance matrix.

4.3 Three full-information methods

4.3.1 Introductory definitions. In this subsection we will define a structural form oriented estimator and two reduced form oriented estimators. One of these is identical to the Three Stage Least Squares (3SLS) method of Zellner and Theil (1962). All three methods stand in the same relation to an estimator in the previous section as does three stage least squares to two stage least squares.

The derivations departure from the structural form of the respecified system which is written for all behavioral relations and all observations simultaneously as

$$\begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_p \end{pmatrix} = \begin{pmatrix} H^*_{(1)}Z_{(1)} & \cdot & \cdot & 0 & 0 & \cdot & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & & 0 & \cdot & \cdot & & H^*_{(p)}Z_{(p)} \end{pmatrix} \begin{pmatrix} \beta'_{(1)} \\ \gamma'_{(1)} \\ \cdot \\ \beta'_{(p)} \\ \gamma'_{(p)} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_p \end{pmatrix} \qquad (1)$$

or more compactly as

$$\text{vecp } Y = \chi\, a + \epsilon \qquad (2)$$

where the p in vecp denotes that the vector is of order Tp. The notations in (1) were given in connection to (4.1.2). In the following we will assume that the covariance matrix of the disturbances is of the form $\Xi_L = \Xi \otimes I_T$, where Ξ is a p x p matrix with typical element $\xi_{ij} = E(\epsilon_{it}\epsilon_{jt})$. This implies that the covariance matrix of the disturbances of the structural form of the CLID-system is of the same form or $E(\delta\, \delta') = \Psi_L = \Psi \otimes I_T$ where Ψ is a p x p matrix. Both covariance matrices are assumed to be nonsingular.[1]

The estimators which we will derive would all be replaced by the generalized least squares method estimating seemingly unrelated regressions suggested by Zellner (1962) if H^* and thus χ were observable. Notations in the definitions that follow are $X^S = \text{diag}(X^S_1, \ldots, X^S_p)$ and $X^R = \text{diag}(X^R_1, \ldots, X^R_p)$. X^S and X^R are χ estimated utilizing the restricted and unrestricted reduced forms respectively, cf. (4.1. 6a–b).

Definition 1. A structural form oriented estimator, the *Full-Information GEID (FI-GEID)* method, is defined as the real solutions a, if any, of the system

$$X^{S\prime}(\Xi^{-1} \otimes I_T)\,\text{vecp } Y = X^{S\prime}(\Xi^{-1} \otimes I_T)\, X^S a \qquad (3a)$$

$$Y^{*\prime} = (I - B)^{-1} C\, Z' \qquad (3b)$$

where Z is the T x m vector of the observations on the predetermined variables.

Comment 1. The system is nonlinear in the unknown elements of β and Γ since Y^* is a function of these elements and a is a vector of them. We distinguish between different real solutions according to definition 4.2.2. The

[1] The assumptions that $(I - \beta)$ and Ψ are nonsingular implies that Ξ is nonsingular.

real solutions belonging to a sequence of solutions that converge in probability to the true parameter values are thus called estimators in the strict sense if any such exist.

Definition 2. A reduced form oriented estimator is defined as the solution a, if any, of the system

$$X^{R\prime}(\Xi^{-1} \otimes I_T) \operatorname{vecp} Y = X^{R\prime}(\Xi^{-1} \otimes I_T)X^R a \qquad (4a)$$

$$Y_i^* = Z(Z'Z)^{-1}Z'Y_i\,, \quad (i = 1, n) \qquad (4b)$$

where Y_i and Y_i^* are T x 1 vectors.

Definition 3. A reduced form oriented estimator, *the three stage least squares method,* is defined as the solution a, if any, of the system

$$X^{R\prime}(\Psi^{-1} \otimes I_T) \operatorname{vecp} Y = X^{R\prime}(\Psi^{-1} \otimes I_T)X^R a \qquad (5a)$$

$$Y_i^* = Z(Z'Z)^{-1}Z'Y_i\,, \quad (i = 1, n) \qquad (5b)$$

Comment 2. In all three definitions the pertaining covariance matrix is specified as known. We will later indicate that the asymptotic properties are unaffected replacing these matrices by consistent estimates.

Comment 3. Starting from the respecified system the choice of the inverse of the covariance matrix of ϵ as weighting matrix is natural. The derivation of the 3SLS estimator usually takes the CLID system as point of departure. Then the choice of the inverse of the covariance matrix of δ is natural. As is well known this choice is fortunate. Under general conditions the resulting estimator is asymptotically efficient.

Comment 4. The estimator in definition 2 was considered by Norlén (1965) under the name the two stage maximum likelihood estimator.

Comment 5. The reason why the URF-steps (4b) and (5b) are performed as ordinary least squares regressions and not as seemingly unrelated regressions with the inverse of the covariance matrix of ϵ as weighting matrix is simple. Due to the form of the covariance matrix we obtain $(\Xi \otimes I_T)^{-1} = \Xi^{-1} \otimes I_T$, using a property of Kronecker products. In this case the regression

$$\operatorname{vecp} Y^* = \operatorname{diag} Z\,[\operatorname{diag} Z'\,(\Xi^{-1} \otimes I_T)\operatorname{diag} Z]^{-1}\operatorname{diag} Z'\,(\Xi^{-1} \otimes I_T)\operatorname{vecp} Y \qquad (6)$$

where diag Z is a block-diagonal matrix of order pT x pm, collapses into ordinary least squares regressions.

Comment 6. The sample sizes required for estimation differ between the structural form oriented estimator and the reduced form oriented estimators; the two latter are more demanding.

Comment 7. The one to one correspondence between REID and CLID systems implies that the estimators although defined for REID systems share the same properties for CLID systems.

4.3.2 A reduced form oriented estimator: The three stage least squares method. In this subsection we will see the advantage with the inverse of the covariance matrix of δ as weighting matrix by deriving the asymptotic distribution of the 3SLS estimator. The results obtained from the rather detailed treatment are later used in the derivation of the asymptotic distribution of the FI-GEID estimator.

As a start we write the i^{th} behavioral relation of the structural form of the CLID system as

$$Y_i = (Y_{(i)}Z_{(i)})(\beta_{(i)}\gamma_{(i)})' + \delta_i, \quad (i = 1, p) \tag{7}$$

Purging the matrix $Y_{(i)}$ from a stochastic component through the URF step we get

$$Y_i = (Y^*_{(i)}Z_{(i)})(\beta_{(i)}\gamma_{(i)})' + \delta_i + e_{(i)}\beta'_{(i)} \tag{8}$$

which is (7) rewritten. The j^{th} column of $Y^*_{(i)}$ is in the URF step obtained as

$$Y^*_j = Z(Z'Z)^{-1}Z'Y_j \tag{9}$$

and the j^{th} column of the matrix $e_{(i)}$ which contains residuals as

$$e_j = Y_j - Y^*_j. \tag{10}$$

Writing the system (1) utilizing (8) we get

$$\text{vecp } Y = X^R a + \delta + (\beta \otimes I_T)e \tag{11}$$

where X^R was defined in (4.1.6b). Carrying through an Aitken estimation of this CLID system—i.e. a generalized least squares step—it seems natural to use the covariance matrix of δ to obtain

$$X^{R\prime}\Psi_L^{-1}\text{vecp } Y = X^{R\prime}\Psi_L^{-1}X^R a + X^{R\prime}\Psi_L^{-1}\delta + X^{R\prime}\Psi_L^{-1}(\beta \otimes I_T)e. \tag{12}$$

where $\Psi_L^{-1} = \Psi^{-1} \otimes I_T$. The circumstance that the ordinary least squares residuals–the p subcolumns of the pT vector e –are uncorrelated with the regressors–the columns of Z–and the form of Ψ_L^{-1} implies that the last term on the right hand side vanishes, and the estimator of a is obtained by substituting for $X^{R\prime}\Psi_L^{-1}\delta$ the zero vector.

We will now show that this is the 3SLS estimator as it usually is given. It is easily realized that the estimator can be rewritten as

$$a - a = (X^{R\prime}\Psi_L^{-1}X^R)^{-1}X^{R\prime}\Psi_L^{-1}\delta \tag{13}$$

using (11).

Developing the ij^{th} submatrix of $X^{R\prime}\Psi_L^{-1}X^R$ we find that

$$\psi^{ij}\begin{pmatrix} Y^{*\prime}_{(i)} \\ Z'_{(i)} \end{pmatrix}(Y^*_{(j)}Z_{(j)}) = \psi^{ij}\begin{pmatrix} Y^{*\prime}_{(i)} \\ Z'_{(i)} \end{pmatrix} Z(Z'Z)^{-1}Z'(Y^*_{(j)}Z_{(j)}) = \tag{14}$$

$$= \psi^{ij}\begin{pmatrix} Y'_{(i)} \\ Z'_{(i)} \end{pmatrix} Z(Z'Z)^{-1}Z'(Y_{(j)}Z_{(j)})$$

where the first equality is due to the fact that the Y_i^*'s are linear combinations of the rows of Z, and the second follows from relation (10), where the residuals obtained from the URF-regression, $e_j (j = 1, n)$, are uncorrelated with the observations on each predetermined variable. Developing $X^{R\prime}\Sigma_L^{-1}\delta$ similarly we get

$$a - a = \begin{pmatrix} \psi^{11}Q_1'Q_1 & \cdots & \psi^{1p}Q_1'Q_p \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \psi^{p1}Q_p'Q_1 & \cdots & \psi^{pp}Q_p'Q_p \end{pmatrix}^{-1} \begin{pmatrix} \Sigma\psi^{1i}\begin{pmatrix} Y'_{(1)} \\ Z'_{(1)} \end{pmatrix} Z(Z'Z)^{-1}Z'\delta_i \\ \cdot \\ \cdot \\ \cdot \\ \Sigma\psi^{pi}\begin{pmatrix} Y'_{(p)} \\ Z'_{(p)} \end{pmatrix} Z(Z'Z)^{-1}Z'\delta_i \end{pmatrix} \tag{15}$$

which is a usual expression for the 3SLS estimator, see Zellner and Theil (1962) or Dhrymes (1970). The notations are $Q_i = U^{-1}Z'(Y_{(i)}Z_{(i)})$ where $UU' = Z'Z$.

The equation (13) is the starting point when showing consistency and deriving the asymptotic distribution of the 3SLS-estimator.[3] We rewrite the equation in the following form

$$T^{1/2}(a - a) = (T^{-1}X^{R\prime}\Psi_L^{-1}X^R)^{-1}T^{-1/2}X^{R\prime}\Psi_L^{-1}\delta \ . \tag{16}$$

[3]We will mainly follow the procedure of Dhrymes (1970).

The question of the asymptotic distribution now resolves into the question of the asymptotic distribution of the last term on the right hand side or $T^{-1/2}X^{R\prime}(\Psi^{-1} \otimes I_T)\delta$, provided the probability limit of the first term on that side exists and is nonstochastic. Interpreting Y^* at time-point t or y_t^* as an estimate of $\eta_t^* = E(y_t|z_t)$ we have, cf. (1),

$$\text{vecp } Y = \text{vecp } H^* + \epsilon = \text{vecp } Y^* + e \qquad (17a\text{–}b)$$

and

$$X^R = \chi + \epsilon_x - e_x \qquad (18)$$

where all matrices are block-diagonal. $X^R = \text{diag}(Y^*_{(1)} Z_{(1)}, ..., Y^*_{(p)} Z_{(p)})$ as before, χ is the corresponding matrix with $Y^*_{(i)}$ replaced by $H^*_{(i)}$ and $\epsilon_x - e_x$ contains the difference for all terms. It follows that

$$T^{-1}X^{R\prime}\Psi_L^{-1}X^R = T^{-1}\chi'\Psi_L^{-1}\chi + \Delta_1$$

where the first term to the right is of order 1 in probability, 0 (1) i.p., and the second $0(T^{-1/2})$ i.p.[4] This means that we have to show that

$$\text{plim } T^{-1}\chi'\Psi_L^{-1}\chi = \text{plim } T^{-1}\chi'(\Psi^{-1} \otimes I_T)\chi \qquad (19)$$

exists and is nonstochastic. The ij^{th} submatrix, cf. (14), is $\psi^{ij}\chi_i'\chi_j$ where χ_i contains the regressors of relation i ,

$$\chi_i = (H^*_{(i)}Z_{(i)}) = ZJ_i \,. \qquad (20a\text{–}b)$$

J_i is a m x $(n_i + m_i)$ matrix containing the URF-parameters defining the n_i H_i^*'s of relation i and selection elements that assure that the m_i predetermined variables of relation i are included e.g.

$$J_i = \begin{pmatrix} & 1 & 0 \\ & 0 & 1 \\ \pi_1'\pi_2' & 0 & 0 \\ & . & . \\ & 0 & 0 \end{pmatrix} \quad , \qquad (21)$$

if relation i of the REID system has H_1^* , H_2^*, Z_1 and Z_2 as regressor vectors. We write

$$\text{plim } T^{-1}\psi^{ij}\chi_i'\chi_j = \psi^{ij}\text{plim } J_i'(T^{-1}Z'Z)J_j \qquad (22)$$

and conclude that the existence of the nonstochastic probability limit of $T^{-1}Z'Z$, which is assumed nonsingular, implies that the same holds for

[4] For the meaning of the term $O(T^{-1/2})$ i.p. cf. Theil (1971, pp. 358, 375).

$$\text{plim}[T^{-1}\chi'(\Psi^{-1} \otimes I_T)\chi]^{-1} \tag{23}$$

provided plim $T^{-1}\chi'(\Psi^{-1} \otimes I_T)\chi$ is nonsingular.

We now concentrate our efforts to the last term on the right hand side of (13). Establishing consistency of the estimator first, we conclude that

$$\text{plim } T^{-1}X^{R\prime}\Psi_L^{-1}\delta = \text{plim } T^{-1}\chi'(\Psi^{-1} \otimes I_T)\delta = 0 \tag{24}$$

using the same arguments as above plus the assumption of the uncorrelatedness of the predetermined variables and the disturbances. This means that

$$\text{plim } a = a \tag{25}$$

and the 3SLS estimator is consistent, cf. (13).

The asymptotic distribution of $T^{1/2}(a - a)$ is determined by the asymptotic distribution of

$$T^{-1/2}X^{R\prime}(\Psi^{-1} \otimes I_T)\delta = T^{-1/2}\begin{pmatrix} \Sigma x_{1t}^{R\prime}\psi^{1\cdot}\delta_t \\ . \\ . \\ . \\ \Sigma x_{pt}^{R\prime}\psi^{p\cdot}\delta_t \end{pmatrix} \tag{26}$$

where δ_t is the p x 1 vector containing the disturbances at time-point t , $\psi^{i\cdot}$ denotes the i^{th} row of Ψ^{-1} and x_{it}^R the t^{th} row of X_i^R.

The i^{th} submatrix of (26) is rewritten as

$$T^{-1/2}\Sigma x_{it}^{R\prime}\psi^{i\cdot}\delta_t = T^{-1/2}\hat{J}_i'\Sigma z_t'\psi^{i\cdot}\delta_t , \tag{27}$$

where $\hat{J}_i$ of order m x $(n_i + m_i)$ contains estimated URF parameters and elements which are either one or zero, cf. (21), and z_t denotes the predetermined variables at time t . The right hand side of (26) now comes out as

$$T^{-1/2}\begin{pmatrix} \hat{J}_1' & . & . & 0 & . & . & 0 \\ 0 & & & & & & \\ . & & & & & & \\ . & & & & & & \\ 0 & . & . & & & & \hat{J}_p' \end{pmatrix}\begin{pmatrix} \Sigma z_t'\psi^{1\cdot}\delta_t \\ . \\ . \\ . \\ \Sigma z_t'\psi^{p\cdot}\delta_t \end{pmatrix} = T^{-1/2}\hat{J}\Sigma v_t , \tag{28}$$

the asymptotic distribution of which is determined by the asymptotic distribution of $T^{-1/2}\Sigma v_t$, since consistency of URF implies that plim $\hat{J}$ exists as nonstochastic.

The general assumptions made for the vector δ_t in subsection 2.1.3, the assumption of normality is not needed, implies that the disturbance vectors

$(\psi^1 \cdot \delta_t, \ldots, \psi^p \cdot \delta_t)$, (t=1, 2,...), are stochastically independent for different t's, have a covariance matrix with desirable properties and expectation zero. The disturbances also have finite moments up to the fourth order. The exogenous variables are as before assumed nonstochastic. All this together with the regularity assumptions made concerning the behavior of the exogenous variables and the dynamic properties of the system, implies that the central limit theorem of Schönfeld (1971B, p. 125) is applicable.[5] We now have a vector $T^{-1/2}\Sigma v_t$ that converges in distribution to a random variable and a matrix, essentially (23), that converges in probability to a constant and conclude by a convergence theorem of Cramér (1946, p. 254) that the asymptotic distribution of the 3SLS estimator is given as

$$T^{1/2}(a - \alpha) \sim N(0, \text{plim } T[\chi'(\Psi^{-1} \otimes I_T)\chi]^{-1}) . \tag{29}$$

We summarize the discussion in

Theorem 1. Under the assumptions 1–10, where however the disturbances need not be identically distributed as in assumption 9, the 3SLS-estimator

$$a = [X^{R\prime}(\Psi^{-1} \otimes I_T)X^R]^{-1} X^{R\prime} (\Psi^{-1} \otimes I_T)\text{vecp } Y \tag{30}$$

is a consistent estimator of a REID-specified system provided $\text{plim}T(X^{R\prime}(\Psi^{-1} \otimes I_T)X^R)^{-1}$ exists as a nonstochastic matrix with finite elements. Adding assumption 11 and assuming that the disturbances are identically distributed the asymptotic distribution is normal and given in (29).

Comment 8. Dhrymes (1970) e.g. gives the corresponding asymptotic distribution $N(0, \text{plim}T[Q'\Psi_L^{-1}Q]^{-1})$ where the ijth submatrix of $Q'\Psi_L^{-1}Q$ is $\psi^{ij}Q_i'Q_j$ of (15). The consistency of the URF-parameters implied by the assumption that URF yields unbiased predictors means that

$$\text{plim } T^{-1}Q_i'Q_j = \text{plim } T^{-1}\hat{J}_i'Z'Z\hat{J}_j = J_i' \text{ plim}(T^{-1}Z'Z)J_j \tag{31}$$

which indeed is plim $T^{-1}\chi_i'\chi_j$. This means that Dhrymes expression is equivalent to (29). The unknown elements of the J-matrices are URF-parameters, cf. (21).

Comment 9. The GEID assumptions given in subsection 2.1.4 allow for certain

[5] The theorem of Schönfeld is somewhat stronger than necessary referring to convergence in quadratic mean rather than in probability as regards the expression for the asymptotic covariance matrix.

correlations between the disturbances and predetermined variables. With these correlations the estimator need not be consistent, cf. (24).

Comment 10. We will in section 4.4 indicate that a consistent estimate of the covariance matrix Ψ does not change the asymptotic distribution.

4.3.3 The asymptotic distribution of the second reduced form oriented estimator. This estimator was given in definition 2 of subsection 4.3.1. The difference vis-a-vis the 3SLS estimator is that the inverse of the covariance matrix of the disturbance vector ϵ and not δ is used as weighting matrix in a generalized least squares step. Using the rewritten structural form of the CLID system (11) it is realized that we can write the estimator as

$$a - \alpha = (X^{R\prime}(\Xi^{-1} \otimes I_T)X^R)^{-1} X^{R\prime}(\Xi^{-1} \otimes I_T)\delta \,. \qquad (32)$$

The derivation of the asymptotic properties now proceeds exactly in the same way as for the 3SLS estimator. We give the results.

Theorem 2. Under the assumptions 1–10, where however the disturbances need not be identically distributed as in assumption 9, the estimator given in definition 2 consistently estimates a REID-specified system provided plim $T[X^{R\prime}(\Xi^{-1} \otimes I_T)X^R]^{-1}$ exists as a nonstochastic matrix with finite elements. Adding assumption 11 and assuming that the disturbances are identically distributed the asymptotic distribution is obtained as

$$T^{1/2}(a - \alpha) \sim N(0,\ \text{plim}\, T[\chi'(\Xi^{-1} \otimes I_T)\chi]^{-1}$$

$$\text{plim}\, T^{-1}\chi'(\Xi^{-1}\Psi\Xi^{-1} \otimes I_T)\chi\ \text{plim}\, T[\chi'(\Xi^{-1} \otimes I_T)\chi]^{-1}) \qquad (33)$$

provided all the plim's have desirable properties.

Comment 11. The case of unknown disturbance covariance matrices will be discussed in section 4.4.

4.3.4 The asymptotic distribution of the FI-GEID estimator. This structural form oriented estimator was given in definition 1. Inserting the one system there (4.3.3b) in the other (4.3.3a), it is realized that we can write the estimator as

$$F(a,M) = 0 \qquad (34)$$

where F is a vector of n_p functions. The vector a is a n_p x 1 vector of the n_p unknown parameters and M a vector of all moments $T^{-1}Y_i'Z_j$, $T^{-1}Z_i'Z_j$

that occur. The covariance matrix $\Xi_L^{-1} = \Xi^{-1} \otimes I_T$ is assumed known and its element do not enter (34) as arguments. The only functions of Y and Z that appear are second order sample moments which, according to the general assumptions made, converge to the corresponding population moments in probability. This means that consistency is established in the same way as for the GEID estimator in section 4.2 earlier. I.e. the use of the implicit function theorem will under certain regularity conditions lead to the existence of a sequence of FI-GEID estimates in the strict sense for a REID-specified system.

Theorem 3. Under the assumptions 1–10, where however the disturbances need not be identically distributed as in assumption 9, a sequence of FI-GEID estimates in the strict sense exists provided some regularity conditions, including the conditions that F in (34) is a continuous function of a and M and possesses continuous first order partial derivatives, are fulfilled.

Comment 12. Under the GEID specification the system (34) is not in general satisfied by the population quantities; the vector of parameters a and the vector of true moments. Writing the true system as vecp $Y = \chi a + \epsilon$ and multiplying by $\chi'(\Xi^{-1} \otimes I_T)$ to the left it is seen that $\chi'(\Xi^{-1} \otimes I_T)\epsilon$ does not in general vanish since the expression involves terms $\chi_i'\epsilon_j$, $i \neq j$. Thus a sequence of FI-GEID estimates in the strict sense need not exist.

The first step in the derivation of the asymptotic distribution is to obtain an expression for $a - a$. Doing that we will follow the approach used by Lyttkens, deriving the asymptotic covariance matrix of the GEID estimator, see e.g. Mosbaek and Wold (1970). The i^{th} relation of the REID system (1) is written as

$$Y_i = (H^*_{(i)}Z_{(i)})(\beta_{(i)}\gamma_{(i)})' + \epsilon_i \tag{35}$$

and the corresponding estimated relation as

$$Y_i = (Y^*_{(i)}Z_{(i)})(b_{(i)}\ c_{(i)})' + e_i \tag{36}$$

Subtracting (36) from (35) we get

$$\epsilon_i - e_i = Y^*_{(i)}(b'_{(i)} - \beta'_{(i)}) + Z_{(i)}(c'_{(i)} - \gamma'_{(i)}) + (Y^*_{(i)} - H^*_{(i)})\beta'_{(i)}. \tag{37}$$

Performing this operation for all relations we obtain

$$\begin{pmatrix} \epsilon_1 - e_1 \\ \cdot \\ \cdot \\ \epsilon_p - e_p \end{pmatrix} = \begin{pmatrix} Y^*_{(1)}Z_{(1)} & \cdot & 0 & 0 & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & \cdot & \cdot & & & Y^*_{(p)}Z_{(p)} \end{pmatrix} \begin{pmatrix} b'_{(1)} - \beta'_{(1)} \\ \cdot \\ \cdot \\ c'_{(p)} - \gamma'_{(p)} \end{pmatrix} + \begin{pmatrix} Y^*_{(1)} - H^*_{(1)} & \cdot & & 0 \\ & & & \\ 0 & \cdot & \cdot & Y^*_{(p)} - H^*_{(p)} \end{pmatrix} \begin{pmatrix} \beta'_{(1)} \\ \cdot \\ \cdot \\ \beta'_{(p)} \end{pmatrix} \tag{38}$$

which is compactly written as

$$\epsilon - e = X^S(a - a) + (\beta \otimes I_T)X_I^S . \tag{39}$$

According to the definition of a REID system and (35) and (36) we have

$$\text{vecp } Y = \text{vecp } H^* + \epsilon = \text{vecp } Y^* + e \tag{40}$$

This implies that (39) can be rewritten as

$$((I - \beta) \otimes I_T)(\epsilon - e) = X^S(a - a) \tag{41}$$

or

$$\epsilon - e = ((I - \beta)^{-1} \otimes I_T)X^S(a - a) . \tag{42}$$

Multiplying by $X^{S\prime}\Sigma_L^{-1}$ where $\Sigma_L^{-1} = \Sigma^{-1} \otimes I_T$ to the left, rearranging and using the fact that

$$X^{S\prime}\Xi_L^{-1}e = 0 \tag{43}$$

which holds according to the definition of the estimator, we get

$$a - a = [X^{S\prime}\Xi_L^{-1}((I - \beta)^{-1} \otimes I_T)X^S]^{-1}X^{S\prime}\Xi_L^{-1}\epsilon . \tag{44}$$

This expression is similar to the starting point when deriving the asymptotic distributions of the RF-oriented estimators earlier. The crucial difference is that the Y^*'s in X^S depend on the estimates of the structural parameters, see (3b). We have however already considered conditions for the consistency of the estimator. It is therefore possible to proceed as in subsection 4.3.2 to obtain the asymptotic distribution of

$$T^{1/2}(a - a) = [T^{-1}X^{S\prime}\Xi_L^{-1}((I - \beta)^{-1} \otimes I_T)X^S]^{-1}T^{-1/2}X^{S\prime}\Xi_L^{-1}\epsilon \tag{45}$$

where a is a root belonging to a sequence of solutions that converges to the true parameter values. The distribution is essentially determined by the asymptotic distribution of $T^{-1/2}X^{S\prime}\Xi_L^{-1}\epsilon$ provided the probability limit of the other term on the right side exists and is nonstochastic. Applying similar arguments as in (17–19) we find that

$$\text{plim } T^{-1}X^{S\prime}\Xi_L^{-1}((I - \beta)^{-1} \otimes I_T)X^S = \text{plim } T^{-1}\chi'(\Xi^{-1}(I - \beta)^{-1} \otimes I_T)\chi \tag{46}$$

The ij^{th} submatrix of the right hand side is $\mu_{ij}\chi_i'\chi_j$ where μ_{ij} is a scalar constant. χ_i represents the "true" regressors of relation i and can be written

$$\chi_i = Z J_i \tag{47}$$

where J_i is a m x $(n_i + m_i)$ matrix containing the restricted reduced form parameters, defining the n_i H^*'s of relation i, and selection elements that assure that the predetermined variables of relation i are included, cf. (20–21). We write

$$\text{plim } T^{-1}\mu_{ij}\chi_i'\chi_j = \text{plim } \mu_{ij}J_i'(T^{-1}Z'Z)J_j \tag{48}$$

and conclude that the existence of the nonstochastic probability limit of $T^{-1}Z'Z$, which is assumed nonsingular, implies that the same holds for

$$\text{plim}[T^{-1}\chi'(\Xi^{-1}(I - B)^{-1} \otimes I_T)\chi]^{-1}, \tag{49}$$

assuming that the inverse exists.

The derivation of the asymptotic distribution of $T^{-1/2}X^S\Xi_L^{-1}\epsilon$ is similar to the derivation of the asymptotic distribution of the 3SLS estimator in subsection 4.3.2. The consistency of the FI-GEID estimator is again utilized, now to show that plim $\hat{J}_i$, where $\hat{J}_i$ depends on elements of B and C, exists as nonstochastic. We give the final result.

Theorem 4. Under the assumptions 1–11 the asymptotic distribution of the FI-GEID estimator is obtained as

$$\begin{aligned} T^{1/2}(a - \alpha) \sim N(0,\ & \text{plim } T[\chi'(\Xi^{-1}(I - B)^{-1} \otimes I_T)\chi]^{-1} \\ & \text{plim } T^{-1}\chi'(\Xi^{-1} \otimes I_T)\chi \text{ plim } T[\chi'(\Xi^{-1}(I - B)^{-1} \otimes I_T)\chi]^{-1'}) \end{aligned} \tag{50}$$

provided the probability limits in (50) exist as nonstochastic matrices with finite elements. It is also required that F in (34) obey certain regularity conditions.

Comment 13. For the distribution in a more general case see below.

4.4 The consequences of an estimated disturbance covariance matrix

Hitherto we have assumed that the disturbance covariance matrices involved are known. This is hardly a realistic assumption. We will in this section give a treatment of the case where the covariance matrix used in the three stage least squares method is replaced by a consistent estimate, and comment upon the consequences for the other two estimators.[1]

[1] Our approach of proving is generally valid only in the absence of lagged endogenous variables, cf. Maddala (1971).

A consistent estimator of the ijth element of Ψ, is

$$\hat{\psi}_{ij} = T^{-1}(Y_i'Y_j - Y_i'A_{ij}Y_j) \tag{1}$$

where

$$A_{ij} = [(Y_{(i)}Z_{(i)})(Q_i'Q_i)^{-1}Q_i'U^{-1}Z']'\,[Y_{(j)}Z_{(j)}(Q_j'Q_j)^{-1}Q_j'U^{-1}Z'] \tag{2}$$

and where in its turn

$$Q_i = U^{-1}Z'(Y_{(i)}Z_{(i)}) \;;\; UU' = Z'Z\,. \tag{3}$$

The relation (1) defines the ijth element of the covariance matrix of the Two Stage Least Squares (2SLS) residuals of the CLID-system. The 2SLS estimator is consistent, and the estimates $\hat{\psi}_{ij}$ differ from the corresponding true valyes by an amount which is of order $T^{-1/2}$ in probability, $O(T^{-1/2})$ i.p.,

The object of our study is the 3SLS estimator a* ;

$$a^* = [X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)X^R]^{-1}X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)\,\text{vecp}\,Y \tag{4}$$

What we wish to show is that the asymptotic distribution of $T^{1/2}(a^* - a)$ is identical to that of $T^{1/2}(a - a)$, where a is the 3SLS estimator when Ψ is known.

Utilizing a convergence theorem saying that convergence in probability to a random variable implies convergence in distribution to the distribution of that variable, cf. Dhrymes (1970, p. 93), we shall show that

$$\text{plim}\; T^{1/2}(a^* - a) - \text{plim}\; T^{1/2}(a - a) = 0\,. \tag{5}$$

Writing out the expression on the left hand side in detail and simplifying we get

$$\begin{aligned}
&\text{plim}\; T[X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)X^R]^{-1}T^{-1/2}X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)\delta \\
&- \text{plim}\; T[X^{R\prime}(\Psi^{-1} \otimes I_T)X^R]^{-1}T^{-1/2}X^{R\prime}(\Psi^{-1} \otimes I_T)\delta \\
&= \text{plim}\; T[X^{R\prime}(\Psi^{-1} \otimes I_T)X^R]^{-1}[\text{plim}\; T^{-1/2}X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)\delta \\
&- \text{plim}\; T^{-1/2}X^{R\prime}(\Psi^{-1} \otimes I_T)\delta]\,,
\end{aligned} \tag{6}$$

using the consistency of $\hat{\Psi}$ and usual regularity conditions. It is the last term to the right that is the interesting one. The two expressions within the brackets are rewritten according to (4.3.26) as

$$\text{plim}\; T^{-1/2}\Sigma x_{1t}^{R\prime}\hat{\psi}^{1\cdot}\delta_t - \text{plim}\; T^{-1/2}\Sigma x_{1t}^{R\prime}\psi^{1\cdot}\delta_t \tag{7}$$

where for simplicity of notation only the first relation of the interdependent system is considered. x_{1t}^R denotes the tth row of X_1^R and $\psi^{1\cdot}$ is the first row

of Ψ^{-1}. The summations go from 1 to T. From (4.3.27) it follows that we need only consider

$$\text{plim } T^{-1/2}\Sigma z_t' \hat{\psi}^{1\cdot}\hat{\delta}_t - \text{plim } T^{-1/2}\Sigma z_t' \psi^{1\cdot}\delta_t \tag{8}$$

or

$$\text{plim } T^{-1/2}\sum_t z_t' \sum_i \hat{\psi}^{1i}\delta_{it} - \text{plim } T^{-1/2}\sum_t z_t' \sum_i \psi^{1i}\delta_{it} \tag{9}$$

where (i=1,p) and ψ^{1i} denotes the $1i^{th}$ element of Ψ^{-1}. By permutating the order of the summations over t and i it is easily realized that the expression equals the zero vector and (5) follows.

Theorem 1. The conclusions of theorem 4.3.1 establishing the asymptotic distribution of the 3SLS estimator remain valid if the covariance matrix Ψ is consistently estimated.

Turning to the asymptotic distribution of the second reduced form oriented estimator which was given in subsection 4.3.3, we see that the covariance matrices of both the structural, i.e. the CLID structural, and the reduced form disturbances are involved. This does not cause any trouble since they are related $\Xi = (I - B)^{-1}\Psi\,(I - B)^{-1'}$. A consistent estimate of Ξ is obtained from the consistent estimate of Ψ , using the 2SLS-residuals, and the consistent 2SLS-estimate of B, say.

Theorem 2. The conclusions of theorem 4.3.2 establishing the asymptotic distribution of the second reduced form oriented estimator remain valid if the covariance matrices Ψ and Ξ are consistently estimated.

The arguments for the structural form oriented estimator are similar and we state

Theorem 3. The conclusions of theorem 4.3.3 establishing the asymptotic distribution of the FI-GEID estimator remain valid if the covariance matrix Ξ is consistently estimated.

A modification of the estimators when the covariance matrices are unknown is to obtain estimates using the residuals given by the methods themselves. A 3SLS method is e.g. defined as the solution a of the following system, cf. the original definition in section 4.3,

$$X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)\,\text{vecp } Y = X^{R\prime}(\hat{\Psi}^{-1} \otimes I_T)\,X^R a \tag{10a}$$

$$Y_i^* = Z(Z'Z)^{-1}Z'Y_i\,, \quad (i=1,n) \tag{10b}$$

$$\hat{\Psi} = T^{-1}\,(Y' - BY' - CZ')(Y' - BY' - CZ')' \tag{10c}$$

where the residuals are calculated for the p behavioral relations. The asymptotic efficiency is not improved by this, as shown already by Madansky (1964). An advantage with the approach is for the FI-GEID estimator, that we avoid the URF completely and can apply the method to smaller samples.

4.5 The gain in efficiency

4.5.1 An analytical discussion. Comparing the asymptotic efficiency of the full-information estimators suggested with the asymptotic efficiency of the corresponding limited-information estimators we utilize the following definition.

Definition 1. Given are two estimators a_1 and a_2 of the parameter vector a where asymptotically

$$T^{1/2}(a_1 - a) \sim N(0, \Sigma_1) \tag{1}$$

$$T^{1/2}(a_2 - a) \sim N(0, \Sigma_2). \tag{2}$$

We say that a_1 is asymptotically efficient relative to a_2 if the matrix $\Sigma_D = \Sigma_2 - \Sigma_1$ is positive semidefinite.

Dhrymes (1970, pp. 216–217) obtains the difference between the covariance matrix Σ_2 of the asymptotic distribution of the 2SLS estimator of the parameters of the entire system and the corresponding 3SLS covariance matrix Σ_3 as

$$\Sigma_2 - \Sigma_3 = \text{plim } T\, Q\, (\Psi \otimes I_T) Q' \tag{3}$$

where in our framework

$$Q = (X^{R\prime}X^R)^{-1}X^{R\prime} - [X^{R\prime}(\Psi^{-1} \otimes I_T)X^R]^{-1}X^{R\prime}(\Psi^{-1} \otimes I_T) \tag{4}$$

Ψ is positive definite and $\Sigma_2 - \Sigma_3$ is at least positive semidefinite. He further shows that the generalized variance of 3SLS is smaller than or equal to the generalized variance of 2SLS, i.e. $\det \Sigma_2 \geqslant \det \Sigma_3$, and that the diagonal elements of Σ_3 are smaller than or equal to those of Σ_2.

There are two cases when there is no gain in efficiency by the 3SLS method as applied to the system as a whole. The first is the obvious situation where it is known that the disturbances of different relations of the CLID structural form are uncorrelated, which means that the two methods are identical. Secondly it has been shown by Zellner and Theil (1962) that if p* of the be-

havioral equations are overidentified while the remaining $p - p^*$ are just identified then the 3SLS estimates of the parameters of the former set are the same as the estimates obtained when the three stage method is confined to the set of overidentified equations.

The two other comparisons: the second reduced form oriented estimator vs. 2SLS and the FI-GEID estimator vs. GEID, are more difficult to perform analytically. Even if we limit the study to the simple two-relational model below, and use the generalized variance comparing GEID and FI-GEID we have to examine a polynomial, in the correlation coefficient of the disturbances of the two relations, of degree six. *A priori* we expect that FI-GEID is more efficient than GEID but as is seen from the following section this is not generally true.

4.5.2. Some illustrations. To get a hunch of the differences between the asymptotic covariance matrices we give the numerical results for some structures. The model considered is a simple two-relational model with four exogenous variables. Two different sets of structural parameters are considered

Set 1

$$B = \begin{pmatrix} & .5 \\ .5 & \end{pmatrix} \qquad \Gamma = \begin{pmatrix} .4243 & .4243 & & \\ & & .4243 & .4243 \end{pmatrix}$$

Set 2

$$B = \begin{pmatrix} & .5 \\ -.5 & \end{pmatrix} \qquad \Gamma = \begin{pmatrix} .5847 & .5847 & & \\ & & .8112 & .8112 \end{pmatrix}$$

The exogenous variables are assumed uncorrelated. For Set 1 the following three correlation structures are specified for the disturbances

Corr 1

$$\Psi = \begin{pmatrix} .09 & \\ & .09 \end{pmatrix}$$

Corr 2

$$\Psi = \begin{pmatrix} .15 & -.075 \\ -.075 & .15 \end{pmatrix}$$

Corr 3

$$\Psi = \begin{pmatrix} .0643 & .03215 \\ .03215 & .0643 \end{pmatrix}$$

and for Set 2 the following three

Corr 1

$$\Psi = \begin{pmatrix} .25 & \\ & .25 \end{pmatrix}$$

Corr 2

$$\Psi = \begin{pmatrix} .25 & .10 \\ .10 & .25 \end{pmatrix}$$

Corr 3

$$\Psi = \begin{pmatrix} .25 & .15 \\ .15 & .25 \end{pmatrix}$$

Combining the structural parameter sets and the correlation structures we get[1] Structures 11, 12 and 13 for Set 1 and Structures 21, 22 and 23 for Set 2. The specifications given are sufficient to calculate the values in table 1 with the addendum that the covariance matrix of the z's is the identity matrix.

As is expected the 3SLS method is always the best. For the structures 13 and 23, are the FI methods better than the two LI methods. The GEID estimator is as good as 3SLS for structure 12. This is due to the symmetry of $(I - \beta)^{-1}$ and the proportionality between the elements of $(I - \beta)^{-1}$ and the elements of the covariance matrix of the disturbances δ_t. For structure 13 on the other hand GEID is rather poor. Table 1 can naturally not be used for any general conclusions except for the confirmation of the existence of structures where one estimator performs well; such as 2SLS for 11 and 21 and GEID for 12.

[1] The first three structures were used by Ågren (1972) in Monte Carlo experiments.

TABLE 1. Comparisons of the diagonal elements of the asymptotic covariance matrices and of the generalized variances. The values are calculated for T = 1 and are multiplied by 10^3.

Parameter	β_{12}	γ_{1i} (i=1, 2)	β_{21}	γ_{2i} (i=3,4)	The generalized variance
Structure 11					
2SLS	141	101	141	101	.00130
GEID	141	113	141	113	.00203
FI-GEID	141	113	141	113	.00203
RF- or 2[2]	141	130	141	130	.00349
3SLS	141	101	141	101	.00130
Structure 12					
2SLS	234	169	234	169	.0156
GEID	234	150	234	150	.0088
FI-GEID	234	169	234	169	.0156
RF- or 2	234	225	234	225	.0479
3SLS	234	150	234	150	.0088
Structure 13					
2SLS	100	72	100	72	.0000970
GEID	100	96	100	96	.000297
FI-GEID	100	67	100	67	.0000685
RF- or 2	100	70	100	70	.0000970
3SLS	100	64	100	64	.0000846
Structure 21					
2SLS	297	266	571	310	.662
GEID	297	297	571	341	1.035
FI-GEID	297	297	571	341	1.035
RF- or 2	297	266	571	310	.662
3SLS	297	266	571	310	.662
Structure 22					
2SLS	297	266	571	310	.467
GEID	297	347	571	291	.656
FI-GEID	297	281	571	305	.502
RF- or 2	297	247	571	296	.349
3SLS	297	246	571	290	.330
Structure 23					
2SLS	297	266	571	310	.271
GEID	297	372	571	266	.326
FI-GEID	297	247	571	273	.161
RF- or 2	297	222	571	281	.135
3SLS	297	221	571	265	.111

[2] This is the second reduced form oriented full-information estimator.

CHAPTER 5: AN EXTENSION OF STRUCTURAL AND REDUCED FORM ORIENTED ESTIMATORS

5.1 Introduction and two principles

The basic interdependent system is here as before

$$y_t = \beta y_t + \Gamma z_t + \delta_t , \quad (t=0, T) \tag{1}$$

where the disturbances are assumed to be generated by the first order autoregressive process

$$\delta_t = R\delta_{t-1} + \delta_t^+ , \quad (t=1,T) . \tag{2}$$

We will discuss methods for estimating the parameters specified in (1) and (2). Within this framework two estimation approaches are considered. They are introduced through the following simple one relational model

$$y_t = \gamma z_t + \delta_t, \quad \delta_t = \rho\delta_{t-1} + \delta_t^+ , \tag{3a–b}$$

where γ and ρ are scalar parameters and y_t, z_t, δ_t and δ_t^+ denote single variables. The notations in (3a–b) apply to the rest of this section only.

Equation (3b), where the δ_t^+'s are serially uncorrelated, specifies a covariance pattern for the disturbances δ . Given that ρ is known we can form the Generalized Least Squares (GLS) estimator of γ ,

$$c = (z'Q^{-1}z)^{-1} z'Q^{-1}y \tag{4}$$

which is a minimum variance linear estimator. Q is the covariance matrix of the error vector δ; z and y denote observation vectors. Normally ρ is unknown but this parameter can be consistently estimated by applying ordinary least squares to (3a), calculating the corresponding residuals and applying ordinary least squares to (3b). Q can now be consistently estimated and through a GLS-step a consistent estimate of γ is obtained. One can show that the asymptotic distribution of the estimator so defined is identical to the distribution of the original GLS estimator (4). The procedure where the autoregressive process is used to consistently estimate a covariance matrix which is used in a generalized least squares step will be referred to as the *GLS approach.*

The transformed system

$$y_t = \gamma z_t + \rho y_{t-1} - \rho\gamma z_{t-1} + \delta_t^+ \tag{5}$$

as obtained from (3a–b) is the base presenting the *calculated residuals as regressors approach* which is the second principle. Rewriting (5) as

$$y_t = \gamma z_t + \rho \delta_{t-1} + \delta_t^+ \tag{6}$$

we have notationally eliminated the nonlinear parameter structure. The new difficulty is that δ_{t-1} is unknown and unobservable. It is however, possible to estimate the variable through a least squares step applied to (3a). Inserting the estimates d_{t-1} in (6) the parameters are estimated regressing y_t on z_t and d_{t-1}. In this way we have defined a two-step estimator.

A modification of the procedure is to define an estimator using the following set of equations

$$\begin{aligned} y'z &= cz'z + rd_l'z \\ y'd_l &= cz'd_l + rd_l'd_l \\ d &= y - cz \end{aligned} \tag{7}$$

where d_l is d lagged one period. The difference is that the estimates of the disturbances and of γ and ρ should simultaneously satisfy a set of equations. The method may be interpreted as an iterated two-step procedure. In other words; the estimate of γ obtained in the second step above is used to calculate new estimates of the disturbances, which as inserted in (6) are used in a least squares step producing estimates of γ and ρ, of which the former is used to re-estimate the disturbances and so forth until convergence is attained.

The three estimation methods presented for the case where the parameter ρ is unknown–"GLS", two-step and (7)–are discussed in Selén (1974) where z_t is assumed stochastic and generated by a process $z_t = \theta z_{t-1} + w_t$. In that paper it is shown, utilizing results of Grether and Maddala (1972) that the two methods using calculated residuals as regressors are, in terms of asymptotic efficiency, inferior to the GLS-approach method and that (7) is superior to the two-step procedure regarding the same property. Also the small sample properties are examined in Selén (1974). It is through Monte Carlo experiments indicated that a ranking based on the asymptotic results essentially holds for finite samples. One interesting finding is that a modification of the method used to introduce the GLS approach above seems to perform better than its origin for finite samples, although the asymptotic properties are identical. The modification is similar to the modification of the two-step method in (7) earlier, i.e. it is required that the estimates of the disturbances used to estimate Q should be obtained utilizing the estimate of γ given by the method.

We have now introduced the two principles and indicated that one is superior. As regards simultaneous equations one would expect that this difference carries over. Discussing the estimation of simultaneous equations we have to account for other factors such as computational problems. The GLS approach implies in its general form, an estimation of the whole system simultaneously and is more complicated. In the calculated residuals as regressors approach it is in principle sufficient to work with one relation at a time. Therefore this principle will be considered. We will discuss the approach before the second approach, which is more promising from a theoretical point of view, is dealt with.

5.2 The calculated residuals as regressors approach

5.2.1 A structural form oriented method. The reformulated transformed interdependent system consisting of p behavioral relations

$$y_t = \beta\eta_t^{**} + R(I - \beta)y_{t-1} + \Gamma z_t - R\Gamma z_{t-1} + \epsilon_t^+ \qquad (1)$$

was introduced in subsection 2.2.1. Simplifying the expression by introducing δ_{t-1} as defined from the CLID system (5.1.1) we obtain

$$y_t = \beta\eta_t^{**} + \Gamma z_t + R\delta_{t-1} + \epsilon_t^+ \qquad (2)$$

which is the base defining the estimator. With (2) it is possible to allow for a non-diagonal autoregressive structure matrix, and still obtain estimates through rather simple procedures as will be discussed in chapter 6. The i^{th} relation of the estimated correspondent to (2) is written as

$$\begin{aligned} Y_i &= (Y_{(i)}^{**} Z_{(i)} d_{(i)l})(b_{(i)} c_{(i)} r_{(i)})' + e_i^+ \\ &= (X_i^{ST} d_{(i)l})(a_i' r_{(i)})' + e_i^+ \end{aligned} \qquad (3a-b)$$

where Y_i is the T x 1 vector $Y_i = (y_{i1},..., y_{iT})'$ and $Y_{(i)}^{**}$, $Z_{(i)}$ and $d_{(i)l}$ are corresponding data matrices of relation i. The symbol X_i^{ST} indicates that the Y^{**}'s are estimated from the structural parameters, through the restricted reduced form. We are now prepared to define a structural form oriented estimator.[1]

Definition 1. The *GEIDA estimator* is defined as the real solutions $a_i r_{(i)}$ (i=1,p), if any, of the system

[1] This estimator was introduced in Selén (1973A–B).

$$(X_i^{ST} d_{(i)l})' Y_i = (X_i^{ST} d_{(i)l})' (X_i^{ST} d_{(i)l})(a_i' r_{(i)})' \ , \ (i=1,p) \qquad (4a)$$

$$Y^{**\prime} = (I - B)^{-1} C\, Z' + (I - B)^{-1} \hat{R} D_l' \qquad (4b)$$

$$d_j = Y_j - Y_{(j)} b_{(j)}' - Z_{(j)} c_{(j)}' \ , \ (j=1,p) \qquad (4c)$$

where the i^{th} column of the T x p matrix D_l is d_{il} .

Comment 1. It is easily realized that the GEIDA estimator is reduced to the GEID estimator if all elements of R are zero *a priori* and if the first observation is included. In that case (4a–b) are identical to (4.2.1a–b) and (4c) is redundant.

Comment 2. Being a highly nonlinear estimator the GEIDA method may give rise to several solutions. By use of the implicit function theorem the existence of an estimator in the strict sense is established, cf. section 4.2.

Comment 3. The complications in the case of identities are minor. Still the p behavioral relations can be written as in (3a–b), which leaves (4a) unchanged. Similarly (4c) remains. The Y_i^{**}'s in (4b) are obtained for all the p + q = n endogenous variables by changing B and C in accordance with (2.1.3) and replacing $\hat{R} D_l'$ by $(D_l \hat{R}' \,|\, 0)'$ where 0 is the zero matrix of order T x q.

5.2.2. The asymptotic distribution. The derivation of the asymptotic distribution of the GEIDA estimator is similar to the corresponding derivations for the structural form oriented methods of chapter 4. The reformulated transformed system

$$y_t = \beta \eta_t^{**} + \Gamma z_t + R\delta_{t-1} + \epsilon_t^+ \qquad (5)$$

is rewritten as

$$y_t = \beta \eta_t^{**} + \Gamma z_t + R d_{t-1} + R(\delta_{t-1} - d_{t-1}) + \epsilon_t^+ \ . \qquad (6)$$

Subtracting the corresponding estimated system

$$y_t = B y_t^{**} + C z_t + \hat{R} d_{t-1} + e_t^+ \qquad (7)$$

we obtain, after adding and subtracting βy_t^{**} on the right hand side, the relation

$$\epsilon_t^+ - e_t^+ = (B - \beta) y_t^{**} + (C - \Gamma) z_t + (\hat{R} - R) d_{t-1} + \beta(y_t^{**} - \eta_t^{**}) - \\ - R(\delta_{t-1} - d_{t-1}) \qquad (8)$$

Developing the difference $y_t^{**} - \eta_t^{**}$ we can write the system for all observations as

$$\begin{pmatrix} \epsilon_1^+ - e_1^+ \\ . \\ . \\ . \\ \epsilon_p^+ - e_p^+ \end{pmatrix} = ((I-\beta)^{-1} \otimes I_T) \begin{pmatrix} Y_{(1)}^{**} Z_{(1)} d_{(1)l} & & 0 \\ . & & \\ . & & \\ . & & \\ 0 & . \; . & . \; Y_{(p)}^{**} Z_{(p)} d_{(p)l} \end{pmatrix} \begin{pmatrix} a_1 - a_1 \\ r'_{(1)} - \rho'_{(1)} \\ . \\ . \\ a_p - a_p \\ r'_{(p)} - \rho'_{(p)} \end{pmatrix}$$

$$- ((I-\beta)^{-1} \otimes I_T) \begin{pmatrix} \delta_{(1)l} - d_{(1)l} & . \; . & . \; 0 \\ . & & \\ . & & \\ 0 & . & . \; \delta_{(p)l} - d_{(p)l} \end{pmatrix} \begin{pmatrix} \rho'_{(1)} \\ . \\ . \\ \rho'_{(p)} \end{pmatrix} \qquad (9)$$

where the notations were explained in connection to (3a–b). In compact notation (9) is written as

$$\epsilon^+ - e^+ = ((I - \beta)^{-1} \otimes I_T) X^{**} (\hat{\theta} - \theta) - ((I - \beta)^{-1} \otimes I_T)(\Delta_{dl} - D_{dl})\rho \qquad (10)$$

where θ is the column vector of the unknown parameters and X^{**} and $\Delta_{dl} - D_{dl}$ are blockdiagonal matrices with pT rows. Multiplying by $X^{**\prime}$ to the left we obtain

$$X^{**\prime}\epsilon^+ = X^{**\prime}((I - \beta)^{-1} \otimes I_T) X^{**} (\hat{\theta} - \theta) - X^{**\prime}((I - \beta)^{-1} \otimes I_T)(\Delta_{dl} - D_{dl})\rho \qquad (11)$$

using a least squares property of the residuals e^+. The structural parameters are the most important quantities and we are satisfied considering these only. Therefore the equations of (11) are permutated to obtain the unknown elements of R in the last positions of θ. The result is

$$(X^{ST} D_{dl})' \epsilon^+ = (X^{ST} D_{dl})' ((I - \beta)^{-1} \otimes I_T) (X^{ST} D_{dl})(a' - a' \;\; r' - \rho')' - (X^{ST} D_{dl})' ((I - \beta)^{-1} \otimes I_T) (\Delta_{dl} - D_{dl})\rho \qquad (12)$$

where $X^{ST} = \text{diag}\,(Y_{(1)}^{**} Z_{(1)}, \ldots, Y_{(p)}^{**} Z_{(p)})$ is the matrix obtained eliminating the columns of X^{**} containing the residuals d_{jl}; these columns form the block-diagonal matrix D_{dl}. Provided the inverse in the following expression exists we obtain

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} = [T^{-1}(X^{ST}D_{dl})'((I-\beta)^{-1} \otimes I_T)(X^{ST}D_{dl})]^{-1}$$

$$[T^{-1/2}(X^{ST}D_{dl})'\epsilon^{+} - T^{-1/2}(X^{ST}D_{dl})'((I-\beta)^{-1} \otimes I_T)$$

$$\cdot(\Delta_{dl} - D_{dl})\rho] . \qquad (13)$$

Let us introduce the notation A for the matrix within the brackets to the left on the right hand side. The T x 1 vector $\delta_{jl} - d_{jl}$ can be rewritten as $(Y_{(j)l}Z_{(j)l})(a_j - \alpha_j)$ which gives

$$(\Delta_{dl} - D_{dl})\rho = X_{pl}R_p(a - \alpha) . \qquad (14)$$

X_{pl} is the blockdiagonal matrix $I_p \otimes (Y_{(1)l}Z_{(1)l}, \ldots, Y_{(p)l}Z_{(p)l})$ and R_p is the $p\Sigma(n_i + m_i)$ x $\Sigma(n_i + m_i)$ matrix where the first $\Sigma(n_i + m_i)$ rows contain a blockdiagonal submatrix with $\rho_{1j}I_{n_j + m_j}$ as its j^{th} block. In case ρ_{1j} is zero this block is the zero matrix. In the following $\Sigma(n_i + m_i)$ rows $\rho_{2j}I_{n_j + m_j}$ forms the j^{th} block and so forth. Now (13) is rewritten as

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} = A^{-1}T^{-1/2}(X^{ST}D_{dl})'\epsilon^{+}$$

$$- A^{-1}T^{-1/2}(X^{ST}D_{dl})'((I-\beta)^{-1} \otimes I_T)X_{pl}R_p(a-\alpha) . \qquad (15)$$

The expression to the right on the right hand side can be developed as

$$A^{-1}T^{-1/2}\begin{pmatrix} X^{ST\prime}((I-\beta)^{-1} \otimes I_T)X_{pl}R_p(a-\alpha) \\ D_{dl}'((I-\beta)^{-1} \otimes I_T)X_{pl}R_p(a-\alpha) \end{pmatrix} =$$

$$= A^{-1}T^{-1/2}\begin{pmatrix} X^{ST\prime}((I-\beta)^{-1} \otimes I_T)X_{pl}R_p & 0 \\ D_{dl}'((I-\beta)^{-1} \otimes I_T)X_{pl}R_p & 0 \end{pmatrix}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} . \qquad (16)$$

Provided the appropriate inverse exists (15) comes out as

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} = \left[I + A^{-1}T^{-1}\begin{pmatrix} X^{ST\prime}((I-\beta)^{-1} \otimes I_T)\ X_{pl}R_p & 0 \\ D_{dl}'((I-\beta)^{-1} \otimes I_T)\ X_{pl}R_p & 0 \end{pmatrix}\right]^{-1}$$

$$A^{-1}T^{-1/2}\begin{pmatrix} X^{ST\prime} \\ D_{dl}' \end{pmatrix}\epsilon^{+} \qquad (17)$$

To simplify we assume that the basic system—the system before the transformation—does not contain any lagged endogenous variables. Conditions for a consistent root were mentioned in the previous subsection and adding the assumption that the appropriate probability limits exist, the matrix within the brackets in (17) can be written

$$I + \begin{pmatrix} (\chi^{Tr\prime}((I-\beta)^{-1} \otimes I_T)\chi^{Tr})^{-1} & 0 \\ 0 & (\Delta_{dl}'((I-\beta)^{-1} \otimes I_T)\Delta_{dl})^{-1} \end{pmatrix}\begin{pmatrix} FR_p & 0 \\ GR_p & 0 \end{pmatrix} \qquad (18)$$

where the meaning of F and G is obvious. The inverse of (18) is easily obtained and a large sample expression for the $\Sigma\,(n_i + m_i)$ first equations of (17) is

$$T^{1/2}(a-\alpha) \approx [(I + (\chi^{Tr\prime}((I-\beta)^{-1} \otimes I_T)\chi^{Tr})\bar{FR}_p^{-1})^{-1}$$

$$\chi^{Tr\prime}((I-\beta)^{-1} \otimes I_T)\chi^{Tr}]T^{-1/2}X^{ST\prime}\epsilon^{+} \qquad (19)$$

In simpler notation (19) is written

$$T^{1/2}(a-\alpha) \approx A_1\, T^{-1/2}X^{ST\prime}\epsilon^{+}. \qquad (20)$$

The asymptotic distribution is essentially determined by the asymptotic behavior of $T^{-1/2}X^{ST\prime}\epsilon^{+}$.

The i^{th} element $(Y_{(i)}^{**}Z_{(i)})$ of the blockdiagonal matrix X^{ST} can be written as $(ZY_lZ_l)\hat{J}_i$ where $\hat{J}_i$ essentially contains functions of the estimates of the structural parameters and R, cf. (4.3.21 and 47). Consistency implies that plim $\hat{J}_i = J_i$ and by arguments similar to those discussing the FI-GEID estimator earlier we obtain the desired distribution.

Theorem 1. Under the assumptions 1–11 as modified in section 2.2 and the assumption that all predetermined variables are purely exogenous, the asymptotic marginal distribution of the strict GEIDA estimator is

$$T^{1/2}(a-\alpha) \sim N(0,\ A_1\ \text{plim}[T^{-1}\chi^{Tr\prime}(\Sigma \otimes I_T)\chi^{Tr}]A_1') \qquad (21)$$

where A_1 is defined as the matrix within the brackets in (19). Σ denotes the covariance matrix of the disturbances of the reduced form of the transformed system. χ^{Tr} is χ^{ST} with the Y^{**}'s replaced by H^{**}'s. The superscript Tr indicates that the systematic part of the endogenous variables of the transformed system is used in χ^{Tr}.

5.2.3 A reduced form oriented method. Extending the 2SLS estimator in a similar way as the GEID estimator in definition 1 earlier we obtain a reduced form oriented method.

Definition 2. A 2SLS type estimator is defined as the real solutions $a_i r_{(i)}$ (i = 1, p), if any, of

$$(X_i^{RT} d_{(i)l})' Y_i = (X_i^{RT} d_{(i)l})' (X_i^{RT} d_{(i)l})(a_i' r_{(i)})', \ (i=1,p) \tag{22a}$$

$$Y_i^{**} = (ZY_l Z_l)\,[(ZY_l Z_l)'\,(ZY_l Z_l)]^{-1} (ZY_l Z_l)' Y_i\,, \ (i=1,p) \tag{22b}$$

$$d_j = Y_j - Y_{(j)} b_{(j)}' - Z_{(j)} c_{(j)}'\,, \ (j=1,p)\,, \tag{22c}$$

where the superscript RT is used to emphasize that the Y_i^{**}'s are obtained from the unrestricted Reduced form corresponding to the Transformed system.

As is seen this nonlinear estimator is demanding as regards the sample size required. In (22b) the endogenous variables are regressed on all predetermined variables and all endogenous and predetermined variables lagged. We expect that the properties of this estimator are inferior to those of the corresponding methods obtained through the GLS-approach. Since the computational properties are not considerably better, we have here no problem with a "genuine" nonlinear estimator, the method will not be considered further.

5.3 The GLS-approach I: Limited-Information methods

5.3.1 Introduction. In this section we will consider modifications of the 2SLS and GEID estimators which account for serial correlation of the disturbances. The starting point is the structural form of the reformulated transformed system (5.2.1) which for all observations is written as

$$\begin{pmatrix} Y_1 - \sum_k \rho_{1k} Y_{kl} \\ \cdot \\ \cdot \\ \cdot \\ Y_p - \sum_k \rho_{pk} Y_{kl} \end{pmatrix} = \begin{pmatrix} \chi_1^{Tr} - \rho_{11} X_{1l}, -\rho_{12} X_{2l} \cdots & -\rho_{1p} X_{pl} \\ \cdot \\ \cdot \\ \cdot \\ -\rho_{p1} X_{1l} & \cdots \chi_p^{Tr} - \rho_{pp} X_{pl} \end{pmatrix}$$

$$\begin{pmatrix} \beta'_{(1)} \\ \cdot \\ \cdot \\ \cdot \\ \gamma'_{(p)} \end{pmatrix} + \begin{pmatrix} \epsilon_1^+ \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_p^+ \end{pmatrix} \qquad (1)$$

where the Y_i's are T x 1-vectors. Y_{il} is Y_i lagged one period, $\chi_i = (H^{**}_{(i)} Z_{(i)})$ and X_{il} is $X_i = (Y_{(i)} Z_{(i)})$ lagged one period. The summation index k ranges from 1 to p. As before Y_i contains the observations on the explained variable of relation i and X_i the observations on the corresponding explanatory variables. The system (1) is complicated with every parameter in β and Γ appearing in every relation, a circumstance which in principle calls for a full-information estimation. To simplify and retain the limited-information feature of the estimators to be considered, we assume that R is diagonal.[1] Now the i^{th} relation can be written

$$Y_i - \rho_{ii} Y_{il} = (\chi_i^{Tr} - \rho_{ii} X_{il}) a_i + \epsilon_i^+ \ , \qquad (2)$$

where a_i is the $n_i + m_i$ column vector of the unknown parameters of relation i. The restricted reduced form of the system (1) is

$$y_t = (I - \beta)^{-1} [\Gamma z_t + R((I - \beta) y_{t-1} - \Gamma z_{t-1})] + \epsilon_t^+ \qquad (3)$$

written at timepoint t.

5.3.2 A reduced form oriented method. We define the estimator directly in the most general case to be considered. Estimates of $H^{**}_{(i)}$ are obtained from the unrestricted reduced form of the transformed system, where z_t, y_{t-1} and z_{t-1} are the regressors. Estimates of ρ_{ii}, denoted r_{ii}, (i=1,p) are obtained from the structural disturbances δ_i, cf. (5.1.2). X_i and d_i denote χ_i and δ_i estimated respectively and the subscript l indicates that the corresponding matrix contains observations on variables lagged one period.

[1] For a suggestion when R is nondiagonal see (1.5.17).

Definition 1. A reduced form oriented estimator, *the A2SLS estimator*, is defined as the real solutions a_i and r_{ii} (i=1,p), if any, of the system

$$(X_i^{RT} - r_{ii}X_{il})'(Y_i - r_{ii}Y_{il}) = (X_i^{RT} - r_{ii}X_{il})'(X_i^{RT} - r_{ii}X_{il})a_i \tag{4a}$$

$$Y_i^{**} = (ZY_lZ_l)\,[(ZY_lZ_l)'(ZY_lZ_l)]^{-1}(ZY_lZ_l)'Y_i \tag{4b}$$

$$d_{il}'d_i = d_{il}'d_{il}r_{ii}\,. \tag{4c}$$

Comment 1. The columns corresponding to the lagged endogenous variables in Z are excluded from (4b) to prevent the nonexistence of the inverse. Corresponding exclusions will also be made in similar expressions in the sequel.

Comment 2. In the case of identities step (4b) is performed for all n endogenous variables.

Comment 3. This estimator is in principle suggested by Fair (1970) and by Dhrymes, Berner and Cummins (1974). In both papers asymptotic covariance matrices are given.

Deriving the asymptotic distribution of the estimator we will first consider the case where R is known. This means that (4a–b) with r_{ii} replaced by the true parameter ρ_{ii} define the estimator.

Carrying through a development similar to that of the 3SLS estimator in subsection 4.3.2 earlier it is seen that the estimator is consistent and that the asymptotic distribution of the i^{th} relation, apart from constants, is determined by the asymptotic distribution of

$$T^{-1/2}\sum_t\left[\begin{pmatrix} y_{(i)t}^{**} \\ z_{(i)t}\end{pmatrix} - \rho_{ii}\begin{pmatrix} y_{(i)t-1} \\ z_{(i)t-1}\end{pmatrix}\right]\delta_{it}^{+} = \hat{J}_iT^{-1/2}\sum_t\begin{pmatrix} z_t \\ z_{t-1} \\ y_{t-1}\end{pmatrix}\delta_{it}^{+} \tag{5}$$

where $\hat{J}_i$ contains ρ_{ii} and the URF parameters defining $y_{(i)t}^{**}$ and selection elements that assure that the $z_{(i)t}$'s are included, cf. (4.3.21). Consistency of the URF means that we need not bother about $\hat{J}_i$.

Applying the central limit theorem of Schönfeld (1971B) we obtain the asymptotic distribution. The joint asymptotic distribution of the vector $a' = (a_1', \ldots, a_p')$ is derived by similar arguments. The result is

$$T^{1/2}(a-\alpha) \sim N(0,\ \text{plim}\ T(Q'Q)^{-1}\text{plim}\ T^{-1}Q'(\Omega \otimes I_T)Q\ \text{plim}\ T(Q'Q)^{-1}) \tag{6}$$

where Q is a blockdiagonal matrix with $\chi_i^{Tr} - \rho_{ii}X_{il}$, as typical element, cf. (2). Ω denotes the covariance matrix of δ_t^+.

In the general case, where R is unknown, is the covariance matrix of formula (6) not valid. The reason is that the estimates a_i and r_{ii} are asymptotica lly correlated. Through the procedure of subsection 5.4.3 below the appropriate expressions are obtained; we give the results.

Theorem 1. Under the assumptions 1–11 as modified is section 2.2 and the assumption that R is diagonal, the asymptotic distribution of the strict A2SLS estimator of definition 5.3.1 is obtained as

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} \sim N(0,\ \text{plim } T \begin{pmatrix} Q'Q & Q'\Delta_l \\ \Delta_l'Q & \Delta_l'\Delta_l \end{pmatrix}^{-1}$$

$$\text{plim } T^{-1}\begin{pmatrix} Q'(\Omega \otimes I_T)Q & Q'(\Omega \otimes I_T)\Delta_l \\ \Delta_l'(\Omega \otimes I_T)Q & \Delta_l'(\Omega \otimes I_T)\Delta_l \end{pmatrix} \text{plim } T \begin{pmatrix} Q'Q & Q'\Delta_l \\ \Delta_l'Q & \Delta_l'\Delta_l \end{pmatrix}^{-1}. \quad (7)$$

where $a' = (a_1', \ldots, a_p')$, $r' = (r_{11}r_{22}, \ldots, r_{pp})$ and $Q = (\chi^{Tr} - (R \otimes I_T)X_l)$. Q is the $pT \times \Sigma(n_i + m_i)$ matrix on the right hand side of (1). Δ_l is the $pT \times p$ blockdiagonal matrix with δ_{il} as its i^{th} block.

5.3.3 A structural form oriented method. We define and state the properties of an estimator similar to the 2SLS estimator of the previous subsection. The difference lies in (8b) below where the restricted reduced form of the transformed system is utilized.

Definition 2. A structural form oriented estimator, *the AGEID estimator* is defined as the real solutions $a_i' = (b_{(i)}c_{(i)})$ and r_{ii} (i=1,p), if any, of the system

$$(X_i^{ST} - r_{ii}X_{il})'(Y_i - r_{ii}Y_{il}) = (X_i^{ST} - r_{ii}X_{il})'(X_i^{ST} - r_{ii}X_{il})a_i \quad (8a)$$

$$Y^{**\prime} = (I - B)^{-1}CZ' + (I - B)^{-1}\hat{R}D_l' \quad (8b)$$

$$d_{il}'d_i = d_{il}'d_{il}r_{ii} \quad (8c)$$

where i=1,p . The i^{th} column of D_l is $d_{il} = Y_{il} - Y_{(i)l}b_{(i)}' - Z_{(i)l}c_{(i)}'$. $X_i^{ST} = (Y_{(i)}^{**}\ Z_{(i)})$.

For the modification in the case of identities see comment 1 below. To

establish consistency of the estimator we can use the approach of chapter 4 subsection 3.4. The asymptotic distribution when R is known is derived as in the previous subsection. We obtain the result

$$T^{1/2}(a-\alpha) \sim N(0, \text{plim } T[Q'((I-B)^{-1} \otimes I_T)Q]^{-1} \text{plim } T^{-1}Q'(\Sigma \otimes I_T)Q$$
$$\cdot \text{plim } T[Q'((I-B)^{-1} \otimes I_T)Q]^{-1\prime}) \qquad (9)$$

where $E(\epsilon^{+}\epsilon^{+\prime}) = \Sigma \otimes I_T$ and Q is a block-diagonal matrix with typical submatrix $\chi_i^{Tr} - \rho_{ii}X_{il}$.

To derive the asymptotic distribution when R is estimated is more complicated. We have to regard all the structural parameters and all the autocorrelation parameters simultaneously. Through the procedure of subsection 5.4.3 below the following result is obtained.

Theorem 2. Under the assumptions 1–11 as modified in section 2.2 and the assumption that R is diagonal, the asymptotic distribution of the strict AGEID estimator of definition 5.3.2 is obtained as

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} \sim N(0, \text{plim } T \begin{pmatrix} Q'((I-B)^{-1} \otimes I_T)Q & Q'\Delta_{dl} \\ \Delta'_{dl}Q & Q'Q \end{pmatrix}^{-1}$$

$$\text{plim } T^{-1} \begin{pmatrix} Q'(\Sigma \otimes I_T)Q & Q'(\Sigma(I-B)^{-1} \otimes I_T)\Delta_{dl} \\ \Delta'_{dl}((I-B)^{-1\prime}\Sigma \otimes I_T)Q & \Delta'_{dl}((I-B)^{-1\prime}\Sigma(I-B)^{-1} \otimes I_T)\Delta_{dl} \end{pmatrix}$$

$$\text{plim } T \begin{pmatrix} Q'((I-B)^{-1\prime} \otimes I_T)Q & \Delta'_{dl}Q \\ Q'\Delta_{dl} & Q'Q \end{pmatrix}^{-1}) \qquad (10)$$

The notations were introduced in theorem 1.

5.4 The GLS approach II: Full-Information methods

5.4.1 R diagonal. The extension of the three full-information methods of chapter 4 to the case where R is diagonal is immediate. Developing the 3SLS estimator, say, in the same way as 2SLS in subsection 5.3.2 above, we get a three stage least squares estimator in the autocorrelated case. For all three full-information methods of chapter 4 the asymptotic normal distributions are obtained through the technique of the following subsections.

5.4.2 R non-diagonal. The system of equations in this general case was written in (5.3.1) which is rewritten compactly as

$$\text{vecp } Y - (R \otimes I_T)\text{vecp } Y_l = (\chi - (R \otimes I_T)X_l)a + \epsilon^+ \quad . \tag{1}$$

The main difference when R is not diagonal is that the matrix of observations on the regressors is not diagonal either, which means that structural parameters of relation i say, are included in other relations. Assuming that the covariance matrix of ϵ^+ is $\Sigma \otimes I_T$, where Σ is a p x p -matrix, it is easily realized that the covariance matrix of δ^+ is of the same form or $\Omega \otimes I_T$. We can use both to define nonlinear estimators and will, assuming that R is unknown, extend the FI-GEID and 3SLS estimators.

Definition 1. A reduced form oriented estimator, *the 3SLSA method,* is defined as the real solutions a and r , if any, of the system

$$(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(\text{vecp } Y - (\hat{R} \otimes I_T)\text{vecp } Y_l) =$$

$$= (X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(X^{RT} - (\hat{R} \otimes I_T)X_l)a \tag{2a}$$

$$Y_i^{**} = (ZY_lZ_l)[(ZY_lZ_l)'(ZY_lZ_l)]^{-1}(ZY_lZ_l)'Y_i \,, \quad (i=1,p) \tag{2b}$$

$$D'_{dl}(\Omega^{-1} \otimes I_T)d = D'_{dl}(\Omega^{-1} \otimes I_T)D_{dl}r \tag{2c}$$

where D_{dl} is a blockdiagonal matrix containing in the i^{th} block the d_i-vectors corresponding to the unknown elements in the i^{th} row of R. The column vector r contains these elements of R row by row. d is the column vector $(d'_1, ..., d'_p)'$. The superscript RT indicates that the Y_i^{**}'s are obtained from the unrestricted reduced form.

Definition 2. A structural form oriented estimator is defined as the real solutions a and r , if any, of the system

$$(X^{ST} - (\hat{R} \otimes I_T)X_l)'(\Sigma^{-1} \otimes I_T)(\text{vecp } Y - (\hat{R} \otimes I_T)\text{vecp } Y_l)$$

$$= (X^{ST} - (\hat{R} \otimes I_T)X_l)'(\Sigma^{-1} \otimes I_T)(X^{ST} - (\hat{R} \otimes I_T)X_l)a \tag{3a}$$

$$Y^{**\prime} = (I - B)^{-1}CZ' + (I - B)^{-1}\hat{R}D'_l \tag{3b}$$

$$D'_{dl}(\tilde{\Omega}^{-1} \otimes I_T)d = D'_{dl}(\tilde{\Omega}^{-1} \otimes I_T)D_{dl}r \tag{3c}$$

where $\tilde{\Omega}^{-1} = (I - B)^{-1\prime}\Sigma^{-1}(I - B)^{-1}$. The superscript ST indicates that

the Y_i^{**}'s are obtained from the restricted reduced form. Other notations are explained in definition 5.3.2 above.

Comment 1. In the case of identities the behavioral relations still are written as in (1) which leaves (2a) and (3a) unchanged. The Y_i^{**}'s are obtained for all n endogenous variables as follows: In (2b) i=1,n and in (3b) $\hat{R}D_l'$ is replaced by $(D_l\hat{R}' \mid 0)'$ where 0 is the zero matrix of order T x q. The covariance matrix of the disturbances of the complete reduced form is of order n; we use the notation Σ_n. The covariance matrix utilized in (3c) is the p x p matrix in the upper left corner of $(I - B)^{-1\prime}\Sigma_n^{-1}(I - B)^{-1}$.

Comment 2. The covariance matrices Ω and Σ are specified as known in definition 1 and 2 respectively. A natural extension is to replace them by consistent estimates. As is the case with the usual 3SLS method we can not normally, when R contains nondiagonal elements, utilize residuals obtained from a 2SLS type method in a "third step". An estimate of the required covariance matrix utilizing the A2SLS residuals is not generally consistent. The trouble is realized by considering (1.5.17). A better approach is to calculate the second order moments of the residuals d_t^+ as obtained from, cf. (1),

$$\text{vecp } Y - (\hat{R} \otimes I_T)\text{vecp } Y_l = (X - (\hat{R} \otimes I_T)X_l)a + d^+$$

where a and $\hat{R}$ or r are given in (2a–c) and X is the blockdiagonal matrix with $(Y_{(i)}Z_{(i)})$ as typical element. The estimator for the general case is defined by adding

$$\hat{\Omega} = T^{-1}\sum_t d_t^+ d_t^{+\prime}$$

to (2a–c) above.

Comment 3. When R is specified as a matrix with no zero-elements *a priori* the Zellner "seemingly unrelated regression steps" in (2c) and (3c) are both reduced to a set of ordinary least squares regressions.

When the elements of R are known, (2c) and (3c) are ignored. Consistency and the asymptotic distributions are in this case established as for the 3SLS and the FI-GEID estimators of chapter 4. We obtain for the 3SLSA method of definition 1

$$T^{1/2}(a-a) \sim N(0, \text{plim}[T^{-1}(\chi^{Tr} - (R \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l)]^{-1} \quad (4)$$

and for the FI-GEID method of definition 2

$$T^{1/2}(a-\alpha) \sim N(0, \text{plim } TA^{-1} \text{plim } T^{-1}F \text{ plim } TA^{-1\prime}) \qquad (5)$$

where

$$A = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Sigma^{-1}(I - \beta)^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l)$$

and

$$F = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Sigma^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l).$$

5.4.3 The asymptotic distribution of the 3SLSA estimator. When there are no restrictions in the covariance matrix of the disturbances, the 3SLS estimator as applied to the basic system, is as efficient as the full-information maximum likelihood method. One would expect that this result carries over to the three stage estimator as applied to the transformed system. This is in fact verified by the quite recent results of Dhrymes and Erlat (1974) who give the estimator and the asymptotic covariance matrix for the case where R is a full matrix. A similar development for our estimator, where some elements of R *a priori* may be zero, goes as follows. We start with the expression for the estimates of the structural parameters.

$$a = [(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(X^{RT} - (\hat{R} \otimes I_T)X_l)]^{-1}$$

$$(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(\text{vecp } Y - (\hat{R} \otimes I_T)\text{vecp } Y_l) \qquad (6)$$

Purging for a stochastic component of Y in the transformed CLID system (2.2.2) utilizing (2b) we obtain

$$\text{vecp } Y = (R \otimes I_T) \text{ vecp } Y_l + (X^{RT} - (R \otimes I_T)X_l)\alpha + (\beta \otimes I_T)e^+ + \delta^+ \quad (7)$$

where e^+ contains the residuals of (2b). With the addition and subtraction of $(\hat{R} \otimes I_T)X_l$ (7) inserted in (6) yields

$$a - \alpha = [(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(X^{RT} - (\hat{R} \otimes I_T)X_l)]^{-1}$$

$$(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)[((R - \hat{R}) \otimes I_T)\delta_l + \delta^+] \qquad (8)$$

Observing that

$$(R \otimes I_T)\delta_l = \Delta_{dl}\rho \ , \qquad (9)$$

where ρ is the column vector containing the non-zero elements of R row by row and Δ_{dl} is a blockdiagonal matrix with the δ_{jl}'s that are regressors in the i^{th} equation of $\delta_t = R\delta_{t-1} + \delta_t^+$ as it's i^{th} block, we can rewrite (8) as

$$a - \alpha = [(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(X^{RT} - (\hat{R} \otimes I_T)X_l)]^{-1}$$

$$(X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(\delta^+ - \Delta_{dl}(r - \rho)) \qquad (10)$$

where r is ρ estimated.
The expression for the estimates of R in definition 1

$$r = (D'_{dl}(\Omega^{-1} \otimes I_T)\ D_{dl})^{-1}\ D'_{dl}(\Omega^{-1} \otimes I_T)d \tag{11}$$

can similarly be developed to

$$r - \rho = (D'_{dl}(\Omega^{-1} \otimes I_T)D_{dl})^{-1}D'_{dl}(\Omega^{-1} \otimes I_T)[\delta^+ - \\ -(X - (R \otimes I_T)X_l)(a - \alpha)]\,. \tag{12}$$

X is defined in comment 2 above. Rewriting (10) and (12) with obvious notations as

$$\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} = \begin{pmatrix} A^{-1}F \\ G^{-1}H \end{pmatrix} \delta^+ - \begin{pmatrix} 0 & A^{-1}F\Delta_{dl} \\ G^{-1}H(X-(R \otimes I_T)X_l) & 0 \end{pmatrix} \begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix}, \tag{13}$$

and multiplying by $\begin{pmatrix} A & 0 \\ 0 & G \end{pmatrix}$ from the left and solving for the parameters we obtain

$$\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} = \begin{pmatrix} A & F\Delta_{dl} \\ H(X-(R \otimes I_T)X_l) & G \end{pmatrix}^{-1} \begin{pmatrix} F \\ H \end{pmatrix} \delta^+ \tag{14}$$

assuming that the inverse exists. Provided the probability limit of T times the first matrix on the right hand side exists as a finite matrix with nonstochastic elements, consistency is established showing that plim $T^{-1}F\delta^+ = 0$ and that plim $T^{-1}H\delta^+ = 0$. It follows from the zero-correlations between the predetermined variables z_t, y_{t-1}, z_{t-1} and the disturbances δ_t^+ that the first probability limit is zero, and from the zero-correlations between δ_{t-1} and δ_t^+ that the second is.

The asymptotic distribution of $T^{1/2}(a' - \alpha'\ r' - \rho')'$ is essentially determined by

$$T^{-1/2}\begin{pmatrix} F \\ H \end{pmatrix}\delta^+ = T^{-1/2} \begin{pmatrix} (X^{RT} - (\hat{R} \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T) \\ D'_{dl}(\Omega^{-1} \otimes I_T) \end{pmatrix}\delta^+ . \tag{15}$$

The elements of the first matrix on the right hand side are developed as follows:

$$X^{RT\prime} = \hat{J}'(I_p \otimes (ZY_lZ_l)') \tag{16}$$

where $\hat{J}$ contains unrestricted reduced form parameter estimates from (2b) and selection elements, cf. (4.3.21),

$$((\hat{R} \otimes I_T)X_l)' = \begin{pmatrix} r_{11}K_1' & \cdots & r_{p1}K_1' \\ \vdots & & \\ r_{1p}K_p' & \cdots & r_{pp}K_p' \end{pmatrix}(I_p \otimes (ZY_lZ_l)') = K'(I_p \otimes (ZY_lZ_l)') \tag{17}$$

where K_i' is a $(m_i + n_i) \times (m + 2n)$ selection matrix. Finally

$$D_{dl}' = \hat{L}'(I_p \otimes (ZY_lZ_l)') \ , \tag{18}$$

where $\hat{L}$ is a matrix which essentially contains structural parameter estimates. The complete right hand side of (15) is written

$$T^{-1/2}\begin{pmatrix} \hat{J}' - K' & 0 \\ 0 & \hat{L}' \end{pmatrix} \quad (I_p \otimes (ZY_lZ_l)')(\Omega^{-1} \otimes I_T)\delta^+ \ . \tag{19}$$

Assuming that the conditions for consistency are fulfilled we only have to rely on the asymptotic normality of $T^{-1/2}(I_p \otimes (ZY_lZ_l)')(\Omega^{-1} \otimes I_T)\delta^+$ to obtain the asymptotic distribution of the estimator, cf. the 3SLS estimator of subsection 4.3.2. Schönfeld's (1971B) central limit theorem is applicable and the distribution in the following theorem is obtained using simplifications valid asymptotically.

Theorem 1. Under the assumptions 1–11 as modified in section 2.2 the asymptotic distribution of the strict 3SLSA estimator of definition 1 is obtained as

$$T^{1/2}\begin{pmatrix} a - \alpha \\ r - \rho \end{pmatrix} \sim N(0 \ , \ \text{plim}\ T\begin{pmatrix} A_1 & Q \\ Q' & U \end{pmatrix}^{-1}) \tag{20}$$

where

$$A_1 = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l)$$

$$Q = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Omega^{-1} \otimes I_T)\Delta_{dl}$$

$$U = \Delta_{dl}'(\Omega^{-1} \otimes I_T)\,\Delta_{dl}$$

5.4.4 The asymptotic distribution of the structural form oriented estimator of definition 2. As with all structural form oriented estimators the derivation of the asymptotic distribution starts with the subtraction of the estimated reformulated system from the true reformulated system. In this case we get, cf. subsection 4.3.4,

$$((I - \beta) \otimes I_T)(\epsilon^+ - e^+) = X^{ST}(a - \alpha) +$$

$$+ ((\hat{R} - R) \otimes I_T)\text{vecp}\ Y_l - (\hat{R} \otimes I_T)X_l a + (R \otimes I_T)X_l\alpha \tag{21}$$

where the notations are obvious from the previous subsections especially formula (1). By adding and subtracting $(\hat{R} \otimes I_T)X_l a$ on the right hand side, rearranging, multiplying all terms by $((I - \beta)^{-1} \otimes I_T)$ from the left and finally multiplying all terms by $(X^{ST} - (\hat{R} \otimes I_T)X_l)'(\Sigma^{-1} \otimes I_T)$ from the left we get

$$\begin{aligned} a - \alpha = &[(X^{ST} - (\hat{R} \otimes I_T)X_l)' (\Sigma^{-1}(I - \beta)^{-1} \otimes I_T) \\ &(X^{ST} - (\hat{R} \otimes I_T)X_l)]^{-1}(X^{ST} - (\hat{R} \otimes I_T)X_l)' \\ &(\Sigma^{-1} \otimes I_T)[\epsilon^+ - ((I - \beta)^{-1}(\hat{R} - R) \otimes I_T)\delta] \end{aligned} \tag{22}$$

which, using (9), in short terms is rewritten as

$$a - \alpha = A_2^{-1}F_1[\epsilon^+ - ((I - \beta)^{-1} \otimes I_T)\Delta_{dl}(r - \rho)] \tag{23}$$

The expression for R in definition 2 is rewritten as, cf. (11) and (12),

$$\begin{aligned} r - \rho = &[D'_{dl}(\tilde{\Omega}^{-1} \otimes I_T)D_{dl}]^{-1} D'_{dl}(\tilde{\Omega}^{-1}(I - \beta) \otimes I_T) \\ &[\epsilon^+ - ((I - \beta)^{-1} \otimes I_T)(X - (R \otimes I_T)X_l)(a - \alpha)] \end{aligned} \tag{24}$$

or in short notation

$$r - \rho = C_1^{-1}H_1[(\epsilon^+ - ((I - \beta)^{-1} \otimes I_T)(X - (R \otimes I_T)X_l)(a - \alpha)] \tag{25}$$

With (23) and (25) we have two expressions that correspond to (13). They are the base establishing the asymptotic distibution.

It is realized, discussing consistency first, that the estimator given in definition 2 can be written as

$$F(a, r, M) = 0 \tag{26}$$

where F is a vector of functions and M a vector of second order moments of elements of y_t, z_t, y_{t-1} and z_{t-1}. This is very similar to results obtained for other structural form oriented estimators earlier, cf. the GEID-estimator (4.2.4). By using the convergence of the sample moments to the corresponding population moments, and imposing some regularity conditions on F, and applying the implicit function theorem consistency is established, see subsection 4.3.4.

Relying on consistency and performing the development of (14) to (19), but with the parameters of the unrestricted reduced form replaced by the parameters of the restricted reduced form, the following asymptotic distribution is obtained.

Theorem 2. Under the assumptions 1–11 as modified in section 2.2 the asymptotic distribution of the strict estimator of definition 2 is obtained as

$$T^{1/2}\begin{pmatrix} a-\alpha \\ r-\rho \end{pmatrix} \sim N(0,\ \operatorname{plim} T \begin{pmatrix} A_3 & F_2 \\ F_3 & G_2 \end{pmatrix}^{-1} \operatorname{plim} T^{-1} \begin{pmatrix} A_4 & F_2 \\ F_2' & G_2 \end{pmatrix} \operatorname{plim} T \begin{pmatrix} A_3' & F_3' \\ F_2' & G_2' \end{pmatrix}^{-1}) \tag{27}$$

where

$$A_3 = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Sigma^{-1}(I-\beta)^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l)$$

$$A_4 = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Sigma^{-1} \otimes I_T)(\chi^{Tr} - (R \otimes I_T)X_l)$$

$$F_2 = (\chi^{Tr} - (R \otimes I_T)X_l)'(\Sigma^{-1}(I-\beta)^{-1} \otimes I_T)'\Delta_{dl}$$

$$F_3 = \Delta_{dl}'((I-\beta)^{-1\prime}\ \Sigma^{-1}(I-\beta)^{-1} \otimes I_T)(X - (R \otimes I_T)X_l)$$

$$G_2 = \Delta_{dl}'((I-\beta)^{-1\prime}\ \Sigma^{-1}(I-\beta)^{-1} \otimes I_T)\Delta_{dl}\,.$$

CHAPTER 6: COMPUTATIONAL METHODS

6.1 Introduction

In the previous chapters we have defined several nonlinear estimators. It is in general not possible to obtain estimates analytically for these methods; numerical algorithms are required. The estimators are given as sets of normal equations, i.e. equation systems to which we wish to find the real roots. There exists a large number of algorithms for this problem.[1] One is the Newton-Raphson method, see (1) below, which is in wide use. A drawback with the Newton-Raphson method is that the derivatives of the normal equations are required for its application. Therefore it often is a laborious task to write computer programs.

For the estimators of chapter 5 and the structural form oriented methods of chapter 4 we will instead use iterative least squares methods. The convergence properties of these methods relative the Newton-Raphson method are not so well explored. The iterative least squares methods are easy to program and have been found to work satisfactory for many problems.

For the FIMLA method of chapter 3 we have given the second order derivatives of the likelihood function. This means that we are a step closer to the application of the Newton-Raphson method as compared to the other estimators. But the expression is complicated and we therefore choose a modification; the Davidon-Fletcher-Powell method.

6.2 The Davidon-Fletcher-Powell method

To obtain the estimates of the FIMLA method we apply a gradient algorithm or, more precisely, the Davidon-Fletcher-Powell (DFP) method, Fletcher and Powell (1963). The DFP method may be termed a quasi-Newton iterative algorithm involving linear searches. The term quasi-Newton means that an attempt is made to simulate the Newton-Raphson iteration

$$\theta^{(s+1)} = \theta^{(s)} - [G^{(s)}]^{-1} g^{(s)} \qquad (1)$$

where $\theta' = (\theta_1, \theta_2, ..., \theta_k)$ is the vector of parameters to be estimated. $g^{(s)}$ is the vector of the first order derivatives and $G^{(s)}$ the matrix of the second

[1] Powell (1971) gives a survey of developments in the field of unconstrained optimization.

order derivatives of minus the likelihood function—or F in (3.2.8)—in the point $\theta = \theta^{(s)}$. The simulation means that a matrix $E^{(s)}$ is used as an estimate of the inverse of $G^{(s)}$. With the notations

$$\theta_D^{(s)} = \theta^{(s+1)} - \theta^{(s)} \qquad (2)$$
$$g_D^{(s)} = g^{(s+1)} - g^{(s)}$$

the matrix $E^{(s+1)}$ is defined as

$$E^{(s+1)} = E^{(s)} + \frac{\theta^{(s)}\theta^{(s)\prime}}{g_D^{(s)\prime}\theta_D^{(s)}} - \frac{E^{(s)}g_D^{(s)}g_D^{(s)\prime}E^{(s)\prime}}{g_D^{(s)\prime}E^{(s)\prime}g_D^{(s)}} \quad . \qquad (3)$$

E is initially any positive definite matrix approximating the inverse of G. In the successive iterations information is built up about the function F so that E finally converges to the inverse of the matrix of the second order derivatives at the minimum.

The computer program written to calculate FIMLA estimates is based on a program by Gruvaeus and Jöreskog (1970) which follows the DFP algorithm with a few modifications. Gruvaeus and Jöreskog have found that their approach is efficient handling problems where (I) the number of arguments of the function is large, (II) the function is complicated and time-consuming to compute, (III) the additional time required to calculate derivative values after the function value has been calculated is relatively small and (IV) second order derivatives are not available or time-consuming to compute. These findings and the property that the algorithm normally is quadratically convergent in the neighbourhood of the solution are promising for the application of the method to our problem of full-information maximum likelihood estimation.

6.3 Fix-point- and related methods

Estimates of the nonlinear estimators of chapters 4 and 5 will be obtained by iterative least squares methods. The main feature of our iterative least squares methods is that one set of parameters is kept constant while the remaining parameters are estimated and vice versa. This idea was considered by Cochrane and Orcutt (1950) who defined an estimator for a single regression equation, with disturbances generated by a first order autoregressive process, cf. (5.1.3). Their method works through a series of least squares steps; on the one hand to obtain estimates of the original parameters of the equation, and on the

other to obtain estimates of the autoregressive parameter. The limited-information methods of section 5.3, AGEID and A2SLS, normally give Cochrane and Orcutt estimates when all the elements of the β-matrix are zero and the interdependent system thus is a number of vector regressions. The idea of the methods of this section will therefore be clear to many readers through the mnemonic Cochrane-Orcutt iterations.

In Wold (1965) the Fix-Point (FP) method was suggested as an iterative method for the estimation of GEID and REID specified systems. This method is the origin of the GEID estimator. We will regard it as a numerical method to calculate GEID-estimates.

The GEID estimator is defined as the solutions to the equations

$$(Y^*_{(i)}Z_{(i)})'Y_i = (Y^*_{(i)}Z_{(i)})'(Y^*_{(i)}Z_{(i)})a_i\,, \quad (i=1,p) \tag{1a}$$

$$Y^{*\prime} = (I - B)^{-1}CZ'\,. \tag{1b}$$

The working principle of the methods of this section is explained as follows. With initial proxies for Y^*, proxies for the structural parameters are determined in (1a). These proxies are inserted in (1b) and new Y^*-values are obtained, which in their turn are used to recalculate proxies for β and Γ. The process is repeated until the differences between the estimates of consecutive iterations are small enough. More formally the estimates of the $(s+1)^{th}$ iteration are obtained in two steps in the FP-method:

Step 1: Proxies of β and Γ, $B^{(s+1)}$ and $C^{(s+1)}$ respectively, are obtained by regressing y_t on $y_t^{(s)}$ and z_t as they occur in each relation giving

$$y_{it} = b^{(s+1)}_{(i)}y^{(s)}_{(i)t} + c^{(s+1)}_{(i)}z_{(i)t} + e_{it}\,, \quad (t=1,T;\ i=1,p) \tag{2a}$$

Step 2: Proxies of η_t^* , $y_t^{(s+1)}$ are for the behavioral relations calculated as

$$y^{(s+1)}_{it} = b^{(s+1)}_{(i)}y^{(s)}_{(i)t} + c^{(s+1)}_{(i)}z_{(i)t},\ (t=1,\ T;\ i=1,p) \tag{2b}$$

and for the identities[1] as

$$y^{(s+1)}_{it} = \beta_{(i)}y^{(s)}_{(i)t} + \gamma_{(i)}z_{(i)t}\,, \quad (t=1,\ T;\ i=p+1,n) \tag{2c}$$

As before the subscript (i) is used to denote that the *a priori* zero parameters with pertaining variables are excluded. Note that (1b) is slightly rewritten in

[1] The original suggestion was that the identities should be eliminated. For a discussion of different approaches reference is made to Ågren (1972).

(2b--c) to get simpler calculations. The residuals e_{it} corresponding to the final estimates B and C are uncorrelated with the z_t's and the final estimates y_t^* occurring in the i^{th} relation. This shows that the FP-method yields GEID estimates. Starting values for the iterative procedure are obtained e.g. by choosing $y_t^{(0)} = y_t$, which has been called OLS-start, or by using the y_t^*'s obtained from the URF, which has been called 2SLS-start.

6.3.1 Improvements of the FP-method. The problem encountered in the FP-method is similar to the problem solving linear equation systems; regard step 2 above. This has enabled a development which has considerably improved the rate and region of convergence for the FP-method.[2] Ågren, see Ågren and Wold (1969) and Ågren (1972), has suggested the Fractional Fix-Point (FFP) method where a relaxation factor is introduced in step 2 above. For those familiar with numerical analysis the method can be understood as Jacobi's method applied together with a relaxation factor. Bodin, see Wold and Lyttkens (1969) and Bodin (1974), has further improved the method by applying the Gauss-Seidel method; his approach can be understood as an extension of the successive overrelaxation method for solving linear equation systems. Since this Recursive Fix-Point (RFP) method is the one we will use some details are given. The matrix B is decomposed into a strictly upper triangular matrix G and a strictly lower triangular matrix H; B = G + H. The two steps above are modified to:

Step 1: Proxies of β and Γ are obtained by least squares as applied to each behavioral relation such that

$$y_{it} = g_{(i)}^{(s+1)} y_{(i)t}^{(s+1)} + h_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t} + e_{it}^{(s+1)} \qquad (3a)$$

where (t=1,T).

Step 2: Proxies of η_t^* (t=1,T) are calculated as

$$y_{it}^{(s+1)} = \mu[g_{(i)}^{(s+1)} y_{(i)t}^{(s+1)} + h_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t}] + (1 - \mu) y_{it}^{(s)} \qquad (3b)$$

For the identities the known parameter values are utilized in step 2. μ is a relaxation factor in the interval (0,2). It should be noted that one first should apply step 1 and thereafter step 2 to the first relation. This is followed by the application of steps 1 and 2 to the relations 2,3 ... respectively.

[2]Improvement of the region of convergence means that the method works for a larger class of structures.

The method is to some extent therefore dependent on the ordering of the relations. The effect of different orderings, and the convergence properties of the RFP method more generally, are extensively studied in Bodin (1974).

6.3.2 The calculation of FI-GEID estimates. The FI-GEID estimator was given in section 4.3. A characteristic feature of the method is that the covariance matrix of the RF disturbances is used in a generalized least squares (GLS) step. In case this matrix is unknown, the estimation procedure starts with the calculation of a consistent estimate. Possible alternatives for that is to utilize the URF-residuals or the 2SLS-residuals, see section 4.4, or the residuals of some other consistent method. We will use the residuals obtained from the REID system estimated by 3SLS or

$$e_t = y_t - By_t^* - Cz_t\,, \quad (t=1,T), \tag{4}$$

unless otherwise is stated. These residuals are used to estimate the ij^{th} element of the covariance matrix Ξ as $T^{-1}\Sigma\, e_{it}e_{jt}$. The estimates of the structural parameters are obtained iterating the following two steps

Step 1: y_t is regressed on $y_t^{(s)}$ and z_t through GLS

$$X^{S(s)'}(\hat{\Xi}^{-1} \otimes I_T)\,\text{vecp}\, Y = X^{S(s)'}(\hat{\Xi}^{-1} \otimes I_T)\, X^{S(s)}\, a^{(s+1)} \tag{5}$$

giving the proxy $a^{(s+1)}$; the column vector that contains the non-zero elements of $B^{(s+1)}$ and $C^{(s+1)}$. In terms of the columns of those matrices we obtain.

$$y_{it} = b_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t} + e_{it}^{(s+1)} \tag{6a}$$

For the notations in (5) see equation (4.3.3a) in the definition of the FI-GEID method.

Step 2: Proxies of η_t^*, $y_t^{(s+1)}$ are calculated as

$$y_{it}^{(s+1)} = \mu[b_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t}] + (1-\mu) y_{it}^{(s)} \tag{6b}$$

where μ is a relaxation factor.

Comment 1. An alternative when the covariance matrix is unknown, is to choose $\hat{\Xi}^{(0)} = I_p$ as an initial estimate and then re-estimate this matrix in each iteration using the residuals of (6a). Or steps 1–2 are iterated with the covariance matrix constant. After convergence Ξ is re-estimated, steps 1–2 are

iterated, when convergence is attained Ξ is estimated anew. We avoid the URF completely and obtain estimates for the modification suggested in section 4.4.

Comment 2. Step 2 is identical to the second step of the FFP method mentioned above.

6.3.3 A fixpoint-method in the serially correlated errors case. For the estimators of chapter 5 the iterative least squares methods become more complicated and the iterations are often performed in more than two steps. We will give procedures for the GEIDA and 3SLSA estimators; for the other methods estimates can be obtained analogously.

The GEIDA method was presented in section 5.2. A fixpoint method, which for the $(s+1)^{th}$ iteration comes out as follows, almost suggests itself.

Step 1: Proxies of B, Γ and R; $B^{(s+1)}$, $C^{(s+1)}$ and $\hat{R}^{(s+1)}$ are obtained by regressing y_t on $y_t^{(s)}$ (the s^{th} proxy of η_t^{**}), z_t and $d_{t-1}^{(s)}$ (the s^{th} proxy of δ_{t-1}) as they occur in each relation giving

$$y_{it} = b_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t} + r_{(i)}^{(s+1)} d_{(i)t-1}^{(s)} + e_{it}^{+(s+1)} \tag{7a}$$

Step 2: Proxies $d_{t-1}^{(s+1)}$ of δ_{t-1}, the structural disturbances of the basic system, are calculated as

$$d_{it}^{(s+1)} = y_{it} - b_{(i)}^{(s+1)} y_{(i)t} - c_{(i)}^{(s+1)} z_{(i)t} \tag{7b}$$

Step 3: Proxies $y_t^{(s+1)}$ of η_t^{**} are calculated as

$$y_{it}^{(s+1)} = b_{(i)}^{(s+1)} y_{(i)t}^{(s)} + c_{(i)}^{(s+1)} z_{(i)t} + r_{(i)}^{(s+1)} d_{(i)t-1}^{(s+1)} \tag{7c}$$

It is assumed that T + 1 observations are available; consequently in all the steps (t=1,T). In practice, to improve the convergence properties, the procedure is modified in analogue with the RFP-method. Thus we first apply all three steps to the first relation obtaining $b_{(1)}^{(s+1)}$, $c_{(1)}^{(s+1)}$, $r_{(1)}^{(s+1)}$, $d_{1t-1}^{(s+1)}$ and $y_{1t}^{(s+1)}$, (t=1,T). These fresh proxies are then used applying the steps to the second relation. We now have $b_{(i)}^{(s+1)}$, $c_{(i)}^{(s+1)}$, $r_{(i)}^{(s+1)}$, $d_{it-1}^{(s+1)}$ and $y_{it}^{(s+1)}$ (t=1,T; i=1,2) when considering the third relation, and the iterative procedure is obvious. A further possible modification is to add a relaxation factor to the third step, cf. (3b).

A choice of starting values is $d_{t-1}^{(0)} = 0$ and $y_t^{(0)} = y_t^{**}$ as determined by the URF, or $y_t^{(0)} = y_t$. We will preferably work with the former $y_t^{(0)}$ alternative.

6.3.4 The calculation of 3SLSA estimates. The 3SLSA estimator of the transformed system was given in subsection 5.4.2. Estimates are obtained by the following procedure.

Step 1: Proxies of β and Γ; $B^{(s+1)}$ and $C^{(s+1)}$ are obtained regressing vecp $Y - (\hat{R}^{(s)} \otimes I_T)$ vecp Y_l on $X^{RT} - (\hat{R}^{(s)} \otimes I_T) X_l$ by GLS giving

$$y_{it} - r_i^{(s)} y_{t-1} = b_i^{(s+1)} y_t^{**} + c_i^{(s+1)} z_t - r_i^{(s)} [B^{(s+1)} y_{t-1} + C^{(s+1)} z_{t-1}] \tag{8a}$$

where $b_i^{(s+1)}$, $c_i^{(s+1)}$ and $r_i^{(s)}$ denote the i^{th} rows of $B^{(s+1)}$, $C^{(s+1)}$ and $\hat{R}^{(s)}$ respectively; (i=1,p).

Step 2: The disturbances of the basic system are calculated as

$$d_{it}^{(s+1)} = y_{it} - b_i^{(s+1)} y_t - c_i^{(s+1)} z_t \ , \quad (i=1,p) \tag{8b}$$

Step 3: Proxies of R; $\hat{R}^{(s+1)}$ are obtained regressing $d_t^{(s+1)}$ on $d_t^{(s)}$ by GLS giving

$$d_{it}^{(s+1)} = r_i^{(s+1)} d_{t-1}^{(s+1)} + d_t^{+(s+1)} \ , \quad (i=1,p). \tag{8c}$$

In step 1 and 3 (t=1,T), whereas in step 2 (t=0,T).

y_t^{**} is, according to the definition, obtained from the unrestricted reduced form, and as starting value for R we choose, unless otherwise is stated, $\hat{R}^{(0)} = 0$.

In the general case where Ω, the covariance matrix of the disturbances δ_t^+ used in step 1 and step 3, is unknown we choose $\hat{\Omega}^{(0)} = I_p$ as starting point. The estimation procedure now is a double iteration method. With $\hat{\Omega}^{(k)}$ as the k^{th} estimate of the covariance matrix, steps 1–3 are iterated. When convergence is attained the $k+1^{th}$ estimate of Ω is calculated utilizing the "final" residuals d_t^+ of step 3. The process is terminated when the successive Ω, β and Γ estimates have converged.

Comment 3. In analogue with the usual 3SLS-estimator we can, in the case where R is diagonal, use the residuals obtained from the A2SLS estimator to estimate the covariance matrix consistently.

Comment 4. The iterative procedure may be improved by introducing a

relaxation factor in step 2 above, thus increasing or decreasing the change in the successive estimates of δ_t. For some specific structures of R we may also improve the performance by modifications in analogue with the RFP-method.

6.4 Multiple solutions

A difficulty with nonlinear estimators is the possibility of multiple solutions; i.e. the existence of several local maxima for maximum likelihood methods and several local minima for least squares methods. The FIMLA estimator of chapter 3 is of the first type whereas the other nonlinear methods can be interpreted as least squares estimators. With a given model, set of data, estimator and iterative method the chosen starting point determines which local extremum is obtained. The calculation of all the extrema requires a very careful and systematic examination of the parameter space for starting values by reproducing the nonlinear surface which determines the estimator. This work is of a magnitude that in practice makes it unrealistic to obtain all the maxima or minima. Possibly one is able to find the extrema where the values of the estimated parameters have economic relevance.

Given several extrema the question arises how to discriminate between them. For maximum likelihood methods the answer is clear; the solution with the highest likelihood should be chosen. For the GEID estimator Edgerton (1973) has suggested that the solution maximizing the sum of the sample R^2-statistics, ΣR_i^2 where

$$R_i^2 = 1 - \Sigma(y_{it}^* - y_{it})^2/\Sigma(y_{it} - \bar{y}_i)^2 , \qquad (1)$$

for all the endogenous variables should be chosen. He shows that this procedure is consistent as regards REID specified systems in the sense that the probability making a wrong choice tends to zero with increasing sample size. For the methods developed for the autoregressive case in chapter 5 we conjecture that a similar procedure is valid. Predictions y_t^{**} for the transformed system are calculated and ΣR_i^2 according to (1) but with y_{it}^* replaced by y_{it}^{**}. For least squares as applied to a single equation we see immediately that this is the correct procedure.

CHAPTER 7: SOME MONTE-CARLO EXPERIMENTS

7.1 Introduction

In this and in the following chapter we will apply many of the methods considered earlier. With the Monte-Carlo experiments we hope to get some indication as to the small sample distribution properties. These are very difficult to extract analytically and a Monte-Carlo study is an alternative which might shed light.

The aspect that we concentrate our interest on concerns (I) the differences between the methods that do not account for serial correlation (standard methods), the methods that account for serial correlation with a diagonal structure matrix R and the methods that account for serial correlation generated by a general first order vector autoregressive process. We hope to indicate a gain in efficiency applying the more sophisticated methods. A second point of interest is (II) a comparison within the groups of methods where R is diagonal and full.

As regards the first point even a small scale sampling experiment will probably be rather informative. In earlier experiments large differences have been demonstrated between some methods accounting for serial correlation and standard methods, see Goldfeld and Quandt (1972) and Selén (1973B). Comparing the methods within the groups where R is diagonal and full, we can utilize the results of the large bulk of sampling experiments comparing methods as 2SLS, FIML, OLS and so on.[1] Our more sophisticated methods are as is obvious from the previous chapters extensions of the standard methods applicable to transformed systems. No results of comparisons within the groups are published and *a priori* we expect that the relative behavior of the extended methods is of similar kind.

In connection with the first aspect above we will (III) examine whether there is a loss in efficiency by applying the methods that account for serial correlation when this complication is absent. A fourth aspect is the validity of asymptotic covariance expressions for small samples (IV). This is a problem area which seems less penetrated and a larger amount of simulations than those following are required for any general conclusions. Through the Monte-Carlo experiments we will also get information as regards the problem of obtaining estimates, i.e. (V) the behavior of the methods used to calculate the estimates.

[1] See the bibliography by Sowey (1973) and the inventory made by Smith (1973). Mosbaek and Wold (1970) perform a large sampling experiment.

7.2 Experimental design

7.2.1 Choice of model. In the Monte-Carlo experiments we will consider the traditional two relational model with four exogenous variables. This model is often referred to as Summers' model as it was used by Summers (1965), and is in its CLID version written as

$$\begin{aligned} y_{1t} &= \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{12} z_{2t} + \delta_{1t} \\ y_{2t} &= \beta_{21} y_{1t} + \gamma_{23} z_{3t} + \gamma_{24} z_{4t} + \delta_{2t} \cdot \end{aligned} \tag{1}$$

The procedure in sampling experiments mainly consists in two steps. Firstly; observations on the variables are generated given a specific structure, i.e. numerical values on the structural parameters, the covariance matrices of the disturbances and the exogenous variables. Secondly; the structural parameters are estimated for each of the generated samples.

In our experiments the disturbances of (1) are generated according to the process

$$\begin{aligned} \delta_{1t} &= \rho_{11} \delta_{1t-1} + \rho_{12} \delta_{2t-1} + \delta^{+}_{1t} \\ \delta_{2t} &= \rho_{21} \delta_{1t-1} + \rho_{22} \delta_{2t-1} + \delta^{+}_{2t} \ , \end{aligned} \tag{2}$$

or the complication for which the estimators are developed. It would also be interesting to examine cases where the disturbances are generated by other processes but this will not be done.

Of the parameters in (1) and (2) the different choices of the elements of R are of primary concern. To satisfy the aims mentioned earlier three cases are important; R full, R diagonal and R empty. We will use the structures

$$\underset{\text{E}}{R = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}, \quad \underset{\text{D1}}{R = \begin{pmatrix} .9 & 0 \\ 0 & .9 \end{pmatrix}}, \quad \underset{\text{F1}}{R = \begin{pmatrix} .6 & .3 \\ .7 & .2 \end{pmatrix}}, \quad \underset{\text{F2}}{R = \begin{pmatrix} .2 & 0 \\ .2 & 0 \end{pmatrix}} \tag{3}$$

Holding all other factors constant, including the covariance matrix of the disturbances δ_t, we are able to study the effects of serial correlation.

Other factors that might be worthwhile considering is the sample size, the variance of the δ_t's or the squares of the multiple correlation coefficients, the covariance of δ_{1t} and δ_{2t}, the correlation of the exogenous variables, the serial correlation of the exogenous variables and the size of the β-coefficients. However, the sampling experiments turned out to be rather costly, due to the large number of complicated methods, and it was decided that

only the sample size should be examined somewhat. For the basic system the sample sizes chosen were 60 and 30 observations. This means that 59 and 29 observations respectively are available for the transformed system.

The structural parameters of the basic system were chosen as

$$B = \begin{pmatrix} .0 & -.5 \\ .5 & .0 \end{pmatrix} \quad \Gamma = \begin{pmatrix} .4243 & .4243 & .0 & .0 \\ .0 & .0 & .4243 & .4243 \end{pmatrix} \tag{4}$$

The covariance matrix of the structural disturbances δ_t was specified as

$$\Psi = \begin{pmatrix} .0643 & \\ .03215 & .0643 \end{pmatrix}. \tag{5}$$

The three matrices in (4) and (5) were used in the example comparing asymptotic covariance expressions in section 4.5. We will use stochastic exogenous variables although most of our theory is developed for the case of fixed z's. The variances of the uncorrelated variables z_{1t}, z_{2t}, z_{3t} and z_{4t} are all 1. As a consequence the squared multiple correlation coefficient of the reduced form of the basic system is 0.8.

The experimental series are

Experiment A	R-structure D1	60 observations
Experiment B	R-structure F1	60 observations
Experiment C	R-structure E	60 observations
Experiment D	R-structure D1	30 observations
Experiment E	R-structure F2	60 observations

The number of replications in each experiment is 50.

7.2.2 Generating samples. Use is made of the restricted reduced form when creating the data. Given Ψ—the covariance matrix of δ_t—and R, Ω—the covariance matrix of δ_t^+—is derived from the relation

$$\Psi = R \Psi R' + \Omega \tag{6}$$

which is a system of linear equations or

$$\text{vec } \Omega = (I - R \otimes R) \text{ vec } \Psi \tag{7}$$

where vec $\Omega = (\omega_{11}\omega_{21}\cdot \omega_{p1}, ..., \omega_{12}\cdot \omega_{1p}\cdot \omega_{pp})'$. Using a slightly modified version of the generator of multivariate normal pseudo random numbers GGNRM in the subroutine library IMSL (1973) we obtained the δ_t^+'s. The observation of δ_t at the first time point was generated also. The subsequent

δ_t observations were calculated utilizing the autoregressive structure (2). The observations of the exogenous variables were generated according to a similar procedure. The observations of the endogenous variables y_t finally were calculated from the restricted reduced form with δ_t, z_t and the parameters β and Γ as known.

7.2.3 Estimation methods. The estimation methods we will consider are given in table 1 with references to the corresponding definitions. It should be noted that the full-information methods 3SLS and FI-GEID and their extensions are modified in comparison to the original definitions. In chapters 4 and 5 the pertaining covariance matrices were initially specified as known to simplify the discussion. As regards the 3SLS estimator we will here follow the usual procedure and use the 2SLS residuals to calculate an estimate of the covariance matrix for the third step. For FI-GEID the 3SLS residuals are used. For 3SLSA the residuals obtained from the structural estimates of that very method are used. The estimator is modified by adding equations for the estimation of Ω, see comment 5.4.2. The equations are

$$\hat{\Omega} = T^{-1}[(I - B)Y' - CZ' - \hat{R}(I - B)Y_l' + \hat{R}CZ_l'] \\ [(I - B)Y' - CZ' - \hat{R}\,(I - B)Y_l' + \hat{R}CZ_l']', \qquad (8)$$

cf. also the discussion in subsection 6.3.4.

TABLE 1. Estimation methods in the sampling experiments.

Group	Notation	Definition	Comment
I	OLS	cf. (1.2.3)	
	2SLS	4.2.3	
	GEID	4.2.1	
	FI-GEID	4.3.1	The covariance matrix in 4.3.1 is estimated, see (7.2.8)
	3SLS	4.3.3	The covariance matrix in 4.3.3 is estimated
	FIML	(3.2.14–15)	Substitute the zero matrix for R
II	A2SLS	5.3.1	
	AGEID	5.3.2	
	3SLSA-Rd	5.4.1	R is a priori specified as diagonal, the covariance matrix in 5.4.1 is estimated.
	FIMLA-Rd	(3.2.14–16)	R is a priori specified as diagonal
III	GEIDA	5.2.1	
	3SLSA	5.4.1	The covariance matrix in 5.4.1 is estimated
	FIMLA	(3.2.14–16)	

7.3 Results

7.3.1 Criteria for judging the outcome. The experiments were conducted on the IBM 370/155 at Uppsala University Data Center using double precision in the floating point arithmetic. Comprehensive results are shown in tables A1–E4 and table 1. The estimate of θ_i, the i^{th} parameter of the model, as obtained by the j^{th} estimation method applied to the k^{th} sample is denoted $\hat{\theta}_{ijk}$. The criteria discussed are as follows:

a) mean square error (MSE) = $K^{-1}\sum_k(\hat{\theta}_{ijk} - \theta_i)^2$, where K is the number of samples

b) mean bias = $K^{-1}\sum_k \hat{\theta}_{ijk} - \theta_i$

c) mean of the standard deviations, or—more precisely—of the square roots of the variances as obtained by the asymptotic formulae, (A) = $K^{-1}\sum_k s(\hat{\theta}_{ijk})$

d) the standard deviations of the empirical sampling distributions; (E) = $(K^{-1}\sum_k(\hat{\theta}_{ijk} - \bar{\hat{\theta}}_{ij})^2)^{1/2}$

e) mean of the R2-statistic as calculated for the CLID structural form (R2 – SF) = $1 - \sum_t d^2_{hkt}/\sum_t(y_{ht} - \bar{y}_h)^2$, where h denotes the h^{th} relation and d_{hkt} the residual

f) mean of the Durbin-Watson "d"-statistic as calculated for the CLID structural form (DW – SF) = $\Sigma(d_{hkt} - d_{hkt-1})^2/\Sigma d^2_{hkt}$

g) mean of the R2-statistic as calculated for the reduced form (R2 – RF) = $1 - \sum_t e^2_{hkt}/\sum_t(y_{ht} - \bar{y}_h)^2$, where e_{hkt} denotes the residual

h) mean of the Durbin-Watson "d"-statistic as calculated for the reduced form (DW – RF) = $\Sigma(e_{hkt} - e_{hkt-1})^2/\Sigma e^2_{hkt}$

i) standard deviations for the R2- and "d"-statistics.

The first four criteria will be discussed in detail. The R2 and "d"-values are as regards the group II and III methods calculated for the transformed system.

7.3.2 The β and Γ parameters. Experiment A where the diagonal elements of R are large can be considered as basic. We discuss the results for the structural parameters as they appear in tables A1–A3. The reader should note that the figures there are multiplied by constants. As indicated in the headings the MSE's are multiplied by 10^5, the mean biases by 10^4 and the "standard deviations" by 10^3. The MSE criterion accounts for both the bias and the variance[1] and as is seen in table A1 the differences in the MSE's

[1] MSE = bias^2 + variance.

for the β's and the γ's are considerable. OLS which is affected both by simultaneity and serial correlation is completely inferior. The differences between the other methods of group I and the methods accounting for serial correlation are also considerable. As regards the β-parameters, for example, the FIMLA methods and the methods obtained through the GLS approach, the latter constitute all the other group II and III methods except GEIDA, have MSE's that are about eight times smaller. GEIDA is, as could be expected, not quite as good with MSE's that are about four times smaller. The differences between the group II and group III methods are small and no loss in estimating the unnecessary ρ_{12} and ρ_{21} parameters is apparent. Within the groups we see that AGEID is inferior to its competitors in group II and that 3SLSA-Rd, as expected, is better than A2SLS.

The standard deviations as obtained from the asymptotic formulae[2] and the empirical sampling distributions are given in table A3. The agreement between the two is, as regards the basic structural parameters, on the whole good. With minor exceptions the asymptotic formulae somewhat understate the variations.[3] This is accentuated for the group I methods. The differences in the standard deviations between the groups are large and it is obvious that much of the differences in the MSE's are attributable to these. For the mean biases in table A2 the picture is not as clear with exception of OLS. The group I methods are however inferior for most parameters.

In experiment B the R-matrix contains one large and one small non-diagonal element. An important purpose of the series is to examine the sensitivity of the group II methods to such misspecification. The MSE's are shown in table B1 and as is seen the group II methods are rather bad for the second relation where the nondiagonal element is large, $\rho_{21} = .7$. As regards β_{21} these methods are worse than the group I methods. Their behavior is hardly superior regarding γ_{23} and γ_{24}.[4] The group III methods on the other hand are good for both relations. It is interesting to note the good performance of GEIDA which is better than A2SLS for 5 out of 6 parameters. Generally though: The differences in the MSE's of this experiment are smaller than those of experiment A. From table B2 it is apparent that the large MSE's of the group II methods partly are due to bias. Table B3 shows that the variances also

[2] No correction for degrees of freedom is made calculating the asymptotic variances.

[3] The asymptotic variances of the maximum likelihood methods are obtained from the E-matrix approximating the inverse of the matrix of the second order derivatives, see section 6.2.

[4] OLS is generally excluded from the following discussion.

contribute. The agreement between the asymptotic and the empirical standard deviations is not as good as in experiment A considering the group II methods. The asymptotic expressions understate the variations more than before. For group I and III the agreement is very good. In contrast to results obtained so far the asymptotic standard deviations of group I are somewhat larger than the empirical standard deviations.

The purpose of experiment C is to examine the loss in efficiency estimating autocorrelation parameters when they are absent. The MSE's are given in table C1. Although there is some loss extending e.g. 3SLS to 3SLSA-Rd first and to 3SLSA later, the differences are minor and their directions are not always unique. For the standard deviations in table C3 the picture is similar to earlier experiments. Some asymptotic standard deviations obtained by using population moments instead of sample moments are given as a complement in the table. They are denoted APM and are apart from constants those of table 4.5.1.

Experiment D is identical to experiment A with the exception that the number of observations is halved. This implies that all the MSE's of table D1 are larger than those of table A1. The differences between the groups in D are similar although not as large. Applying the GEIDA estimator in experiment D we excluded the *a priori* zero elements ρ_{12} and ρ_{21} from the estimation. The relative behavior of the method is hardly improved by this. The agreement between the empirical and the asymptotic standard deviations has decreased as is seen in table D3. The understatements of the variations resulting from the asymptotic formulae are a bit larger than earlier.

In experiment E are two small non-zero elements, ρ_{11} and ρ_{21}, introduced. The outcome of this intermediate experiment is given in tables E1–E3. The differences between the groups are on the whole small but the group II methods seem to be somewhat better than their competitors. The superiority versus 3SLSA is probably due the fact that the restrictions $\rho_{12} = \rho_{22} = 0$ are deliberately ignored in applying that method. This is confirmed by the behavior of FIMLA when the restrictions are taken care of. In experiment B, with a large nondiagonal element in R, bias in β_{21} was indicated for the group II methods. The results in table E2 do not point in that direction. As regards the standard deviations the results are more inconclusive than earlier; sometimes the empirical standard deviations are larger than the asymptotic.

A difficulty in discussing the outcome of the experiments which we have not mentioned is the problem of existence, or rather of non-existence, of the second order moments. For Summers' model it is known that the

variances of the GEID and 2SLS estimators e.g. do not exist. With this in mind the choice of Summers' model perhaps seems bad; extreme results of one single sample may distort the comparison of the mean square errors. Some guard has to be taken against the problem. A visual inspection of the estimates did not reveal any extreme results, and there is no sign of them in the mean biases and the mean square errors given in the tables. This is in accordance with Ågren (1972, p. 131) who applied 2SLS and GEID to the structure defined by our β and Γ parameter matrices and our covariance matrices.

7.3.3 The R parameters. Turning to the R parameters in the basic experiment A we see immediately in table A1, that all MSE's are considerably larger than those of the structural parameters. This is due to both increased biases and increased variances. Neither the biases nor their signs are unexpected. The estimators of the autocorrelation parameter are generally biased already in the simplest one-relational case

$$y_t = \gamma z_t + \delta_t \ ; \ \delta_t = \rho \delta_{t-1} + \delta_t^+ \qquad (1)$$

With true positive values on the scalars γ and ρ the bias is downwards, cf. Rao and Griliches (1969). Comparing the methods of group II and III we see that the MSE's for ρ_{11} and ρ_{22} increase when the *a priori* zero elements are estimated. FIMLA is the worst method.

The agreement between the standard deviations as obtained from the asymptotic formulae and the empirical sampling distributions is not quite as good as for the other parameters.

From table A3 it should further be noted that the standard deviations of the group III methods are larger than those of group II.

The MSE's of ρ_{22} are quite large in experiment B as regards the group II methods. This is due to the biases. The estimates are strongly biased upwards in an attempt to compensate—so it seems—for the large non-diagonal element ρ_{21}. The asymptotic standard deviations of the group III methods written in table B3 are much closer to their empirical correspondents than in table A3. The behavior of the group II methods on the other hand is, not unexpectedly, bad.

The standard deviations are large in experiment C, which has zero R-elements. This gives large MSE's despite the small biases. A decrease in the number of observations in experiment A results in increased biases and a still worse agreement between the asymptotic and the empirical standard deviations; compare the tables A2, A3, D2 and D3. For experiment E, where

the R-elements are small, we find in table E3 that the empirical standard deviations sometimes are larger than their empirical correspondents. This is in contrast to the general impression. As regards all the experiments the FIMLA methods are at the bottom.

7.3.4 Further results and comments. The serial correlation of the disturbances and the goodness of fit are, as regards the CLID specified structural relations, indicated in tables A4–E4. The numbers within the parentheses are the associated standard deviations. The tables illustrate the gain in fit introducing the parameters of R in the estimation. Of more interest are the calculated "d"-values. For the case of three explanatory variables in a single regression we reproduce the critical regions given by Durbin and Watson (1951) in table 1.

TABLE 1: Lower and upper bounds of the Durbin-Watson statistic.

	5 % level		1 % level	
	d_l	d_u	d_l	d_u
60 observations	1.48	1.69	1.32	1.52
30 observations	1.20	1.62	1.01	1.42

The tables A4–E4 contain mean values of the calculated "d"-values. Although the proper statistic is the frequency of "d"-values in the critical regions the tables illustrate the serial correlation of the disturbances. The mean values of the group I methods fall in the critical region except for the experiments C and E where the elements of R are small. As regards the methods of group II and III, the "d"-values are calculated for the transformed system. The mean values are very close to 2 with some exception of experiment B. For a linear model the values indicate absence of serial correlation. A further illustration of the outcome is obtained from table I where some R2- and "d"-values for the reduced form are given.

The behavior of the numerical methods was good in all the experiments. In no sample difficulties with convergence occurred. The convergence of the iterative least squares methods was rapid; the maximum number of iterations required for the A2SLS, GEIDA and 3SLSA methods were, in experiment B e.g., 17. The CPU-time required to generate data, calculate OLS, 2SLS, A2SLS, GEIDA and 3SLSA estimates and associated statistics was about 6 minutes and 30 seconds in that experiment. The calculation of maximum likelihood estimates was slower and a CPU-time of 3 minutes and 30 seconds was required for FIMLA and FIMLA-Rd together. In experi-

ment E the calculation of the OLS, GEID, A2SLS, GEIDA and 3SLSA-Rd estimates lasted 6 minutes and 30 seconds, while FIMLA and FIMLA-Rd required 6 minutes. It should be emphasized that the time comparisons only are approximative since different convergence criteria were used in the two cases. For the iterative least squares methods the testing was performed on the relative differences of the parameter estimates in subsequent iterations.[5] As soon as the inequality

$$\max_i (\theta_i^{(s)} - \theta_i^{(s-1)})/\theta_i^{(s)} < 10^{-6} \quad , \tag{2}$$

where $\theta_i^{(s)}$ denotes the estimate of the i^{th} parameter as obtained in the s^{th} iteration, was satisfied the estimation procedure was terminated. As regards the Davidon-Fletcher-Powell method the testing was performed on the values of the first order derivatives. The iterative procedure was terminated when the inequality

$$\max_i (d\log L/d\theta_i)_{\theta_i = \theta_i^{(s)}} < 5^{-5} \operatorname{tr} \sum_t (y_t - \bar{y})(y_t - \bar{y})' \tag{3}$$

was satisfied. log L denotes the log likelihood.

7.3.5 Conclusions and summary. The results are obtained for Summers' model given one particular β and Γ structure. Nevertheless some important conclusions can be made. When the serial correlation of the disturbances is large there may be a considerable loss in efficiency estimating the structural parameters by the usual methods. For the case where R has large diagonal elements we have obtained big differences between, on the one hand, the group I methods and, on the other, the more sophisticated group II and III methods specified in table 7.2.1. It is further indicated that methods relying on a diagonal R-matrix are inferior to those of group III, they may even be inferior to the group I methods, when a non-diagonal R-element is large. For cases with small or zero R-elements the choice between the groups seems less important. In the former case the group I methods may work well, whilst in the second the loss in estimating the known zero-elements is small; at least when the number of such elements is small as here.

A discrimination of the methods within the groups is difficult. From numerous Monte Carlo experiments we know that the behavior of the different group I methods, except OLS which regularly is inferior, depends on the

[5] This procedure is simple but may occasionally go wrong, see Ågren (1972) and Bodin (1974). The proper testing should be performed on the y^*- or y^{**}-values in subsequent iterations.

specific β and Γ structure chosen. Here that structure is fixed whilst the R-structure is varied. As is seen, the behavior of the methods within the groups vary even with the different choices of R. What we yet can say when discussing efficiency is that the "calculated residuals as regressors" approach method, GEIDA, seems inferior to the other group III methods and that 3SLSA-Rd and FIMLA-Rd probably behave better than A2SLS and AGEID in the case of a not non-diagonal R-matrix.

Goldfeld and Quandt (1972) compared in a similar sampling experiment the behavior of the OLS, 2SLS, FIML and FIMLA-Rd estimators. The methods were applied to a two-relational model with a diagonal R-matrix. Goldfeld and Quandt also indicate a large impact in terms of efficiency loss when the serial correlation is large. Increasing the sample size from 30 to 180 they found a steady improvement for FIMLA-Rd as compared to the other methods. Different sizes of the parameters and different correlations between the exogenous variables were tried but the given conclusions regarding efficiency loss remained valid.

TABLE A1. Mean square errors experiment A. $10^5(K^{-1}\sum_k(\hat{\theta}_{ijk} - \theta_i)^2)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	838	927	243	145	128	163				
2SLS	163	156	163	117	149	115				
GEID	165	155	216	151	178	148				
FI-GEID	173	175	125	101	149	125				
3SLS	164	161	134	100	138	113				
FIML										
A2SLS	21	19	16	17	13	14	682			501
AGEID	23	18	24	25	19	19	689			507
3SLSA-Rd	20	18	13	14	11	13	463			507
FIMLA-Rd	21	18	13	14	11	13	901			1094
GEIDA	45	40	51	32	28	34	1446	904	820	1042
3SLSA	20	19	13	14	11	13	1136	845	798	769
FIMLA	21	19	13	14	11	13	1960	1104	1314	1614

TABLE A2. Mean biases experiment A. $10^4(K^{-1}\sum_k\hat{\theta}_{ijk} - \theta_i)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	754	845	−274	−217	−171	−237				
2SLS	−88	−41	−32	13	68	20				
GEID	−84	−42	−26	−1	83	−3				
FI-GEID	−114	−74	−30	15	48	56				
3SLS	−91	−45	−27	19	54	44				
FIML										
A2SLS	−7	31	−9	4	−17	−23	−373			−260
AGEID	−21	19	−8	−7	−17	−16	−381			−277
3SLSA-Rd	−11	25	−5	8	−7	−19	−294			−196
FIMLA-Rd	−24	13	−5	9	−9	−20	−633			−557
GEIDA	−37	31	0	15	−5	−18	−702	39	−37	−496
3SLSA	−15	23	−4	10	−8	−18	−564	−42	−135	−339
FIMLA	−22	12	−5	−14	8	16	−911	−109	−235	−712

TABLE A3. Square roots of the variances experiment A. The values are multiplied by 10^3.

A = according to the asymptotic formulae; the entries are means calculated over the 50 generated samples.

E = as obtained from the empirical sampling distributions.

		β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	A	31	32	30	29	30	30				
	E	52	47	41	32	32	33				
2SLS	A	38	39	32	32	32	32				
	E	39	39	40	34	38	34				
GEID	A	38	40	37	36	36	37				
	E	40	40	42	42	42	39				
FI-GEID	A	38	40	31	31	31	32				
	E	40	42	36	32	39	35				
3SLS	A	38	39	30	30	30	30				
	E	39	40	37	32	37	33				
FIML	A										
	E										
A2SLS	A	14	14	11	11	11	11	58			57
	E	15	13	13	13	11	12	74			66
AGEID	A										
	E	15	13	16	16	14	14	74			66
3SLSA-Rd	A	13	14	11	11	11	11	56			63
	E	14	13	12	12	10	11	62			68
FIMLA-Rd	A	14	14	11	11	11	11	64			63
	E	15	13	12	12	21	11	71			89
GEIDA	A										
	E	21	20	23	18	19	19	98	95	91	89
3SLSA	A	13	14	11	11	11	11	74	74	75	74
	E	14	14	12	12	10	11	91	93	89	82
FIMLA	A	14	14	11	11	11	11	83	84	84	84
	E	14	14	12	12	10	11	107	106	113	106

TABLE A4. Mean R2- and "d"-values of the CLID specified structural form. Associated standard deviations are given within the parentheses.

	R2 relation 1		R2 relation 2		"d" relation 1		"d" relation 2	
OLS	.945	(.024)	.945	(.027)	.51	(.187)	.50	(.183)
2SLS	.938	(.030)	.938	(.037)	.36	(.152)	.34	(.150)
GEID	.937	(.032)	.936	(.039)	.38	(.155)	.37	(.144)
FI-GEID	.936	(.030)	.935	(.037)	.35	(.156)	.34	(.174)
3SLS	.937	(.030)	.936	(.037)	.35	(.156)	.34	(.143)
FIML								
A2SLS	.987	(.003)	.987	(.004)	1.94	(.221)	1.95	(.226)
AGEID	.987	(.003)	.987	(.004)	1.94	(.222)	1.95	(.228)
3SLSA-Rd	.987	(.003)	.987	(.004)	1.96	(.229)	1.96	(.222)
FIMLA-Rd	.987	(.003)	.987	(.004)	1.89	(.232)	1.85	(.281)
GEIDA	.986	(.003)	.986	(.004)	2.00	(.228)	1.99	(.229)
3SLSA	.987	(.003)	.987	(.004)	1.96	(.218)	1.96	(.220)
FIMLA	.987	(.003)	.986	(.004)	1.85	(.239)	1.82	(.314)

TABLE B1. Mean square errors experiment B. $10^5(K^{-1}\underset{k}{\Sigma}(\hat{\theta}_{ijk} - \theta_i)^2)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1277	1239	165	145	174	195				
2SLS	147	143	97	91	138	108				
GEID	149	141	127	119	198	166				
FI-GEID	149	144	109	88	109	95				
3SLS	144	145	95	82	111	99				
FIML										
A2SLS	94	187	62	52	123	101	5517			14080
AGEID	88	172	44	34	119	89	5249			14410
3SLSA-Rd	85	150	49	38	96	81	18230			2084
FIMLA-Rd	82	149	47	37	98	75	14132			14640
GEIDA	91	64	64	47	58	56	1348	925	997	949
3SLSA	52	55	30	29	48	53	1300	879	1058	977
FIMLA	55	59	29	29	51	55	4723	1474	1381	1092

TABLE B2. Mean biases experiment B. $10^4(K^{-1}\underset{k}{\Sigma}\hat{\theta}_{ijk} - \theta_i)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	997	979	−275	−241	−252	−294				
2SLS	2	32	1	30	−9	−42				
GEID	2	31	20	5	0	−70				
FI-GEID	−36	8	−28	66	−48	2				
3SLS	4	34	−15	50	−29	−10				
FIML										
A2SLS	−48	209	−33	45	−74	−23	1575			2974
AGEID	−48	150	−45	8	−27	−6	1570			2969
3SLSA-Rd	−29	145	−58	5	−25	−8	4064			4279
FIMLA-Rd	−29	154	−60	51	−21	−16	3444			3399
GEIDA	−5	1	−46	35	−13	−7	−355	−198	7	106
3SLSA	38	−3	−25	45	−13	−5	−379	−150	150	76
FIMLA	30	−8	−23	49	−12	−3	−1851	599	220	47

TABLE B3. Square roots of the variances experiment B. The values are multiplied by 10^3.

A = according to the asymptotic formulae; the entries are means calculated over the 50 generated samples.

E = as obtained from the empirical sampling distributions.

		β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	A	33	33	31	30	32	31				
	E	54	54	30	30	34	33				
2SLS	A	41	40	34	34	35	34				
	E	39	38	32	30	37	33				
GEID	A	41	40	39	39	40	39				
	E	39	38	36	35	45	41				
FI-GEID	A	41	41	32	32	34	34				
	E	39	38	33	29	33	31				
3SLS	A	41	40	32	32	33	32				
	E	38	38	31	29	34	32				
FIML	A										
	E										
A2SLS	A	24	29	21	21	26	25	88			102
	E	31	38	25	22	35	32	176			231
AGEID	A										
	E	30	39	21	19	35	30	169			239
3SLSA-Rd	A	23	28	18	18	22	22	85			93
	E	29	36	22	20	31	29	132			159
FIMLA-Rd	A	23	29	18	19	23	23	119			110
	E	29	36	21	19	32	28	151			176
GEIDA	A										
	E	30	26	25	22	24	24	112	95	101	98
3SLSA	A	20	23	16	16	18	18	103	97	94	91
	E	23	24	17	17	22	23	109	93	103	100
FIMLA	A	20	23	16	16	19	19	111	102	100	94
	E	24	25	17	17	23	24	114	106	115	104

TABLE B4. Mean R2- and "d"-values of the CLID-specified structural form. Associated standard deviations are given within the parentheses.

	R2 relation 1		R2 relation 2		"d" relation 1		"d" relation 2	
OLS	.944	(.017)	.941	(.018)	1.14	(.248)	1.24	(.294)
2SLS	.933	(.026)	.931	(.026)	.78	(.336)	.34	(.389)
GEID	.933	(.026)	.930	(.027)	.80	(.335)	1.05	(.105)
FI-GEID	.932	(.028)	.929	(.027)	.75	(.340)	.98	(.381)
3SLS FIML	.933	(.026)	.930	(.026)	.76	(.341)	1.00	(.397)
A2SLS	.966	(.008)	.955	(.012)	2.32	(.162)	2.26	(.204)
AGEID	.966	(.008)	.955	(.012)	2.33	(.164)	2.26	(.210)
3SLSA-Rd FIMLA-Rd	.961	(.011)	.953	(.013)	2.70	(.205)	2.54	(.231)
GEIDA	.973	(.005)	.977	(.006)	1.95	(.108)	1.92	(.160)
3SLSA FIMLA	.973	(.006)	.978	(.006)	1.94	(.108)	1.92	(.160)

TABLE C1. Mean square errors experiment C. $10^5(K^{-1}\sum_k(\hat{\theta}_{ijk} - \theta_i)^2)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1161	1123	189	225	283	185				
2SLS	141	240	131	148	157	159				
GEID	141	246	182	229	236	213				
FI-GEID	157	243	124	106	120	134				
3SLS	144	237	117	114	121	135				
FIML	160	237	122	109	121	137				
A2SLS	151	282	125	144	168	160	1777			2711
AGEID	172	309	181	224	257	209	1776			2683
3SLSA-Rd FIMLA-Rd	150	265	114	121	119	135	1671			2363
GEIDA	161	270	183	226	237	208	1917	2159	2200	2468
3SLSA	162	280	120	124	124	137	2168	2121	2676	2854
FIMLA	182	302	130	123	133	153	2352	2170	2824	2746

TABLE C2. Mean biases experiment C. $10^4(K^{-1}\sum_k\hat{\theta}_{ijk} - \theta_i)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1009	977	−302	−322	−389	−265				
2SLS	−10	−30	−3	−39	−110	18				
GEID	−14	−20	36	−91	−13	8				
FI-GEID	−53	−84	−46	20	−76	12				
3SLS	−7	−40	−25	−5	−87	24				
FIML	−48	−72	−23	−3	−90	27				
A2SLS	106	92	−37	−85	−143	−11	36			−37
AGEID	−16	−6	40	−116	−136	15	10			−37
3SLSA-Rd FIMLA-Rd	68	38	−44	−12	−91	24	−11			−11
GEIDA	16	−6	18	23	−143	−2	−89	24	20	−81
3SLSA	31	21	−33	−10	−98	18	−11	201	211	−129
FIMLA	−56	−81	−33	−7	−89	25	−170	214	244	−306

TABLE C3. Square roots of the variances experiment C. The values are multiplied by 10^3.

A = according to the asymptotic formulae, the entries are means calculated over the 50 generated samples.

E = as obtained from the empirical sampling distributions.

APM = according to the asymptotic formulae utilizing the population moments.

		β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	A	33	33	32	31	32	32				
	E	28	42	32	35	37	34				
2SLS	APM	41	41	35	35	35	35				
	A	41	41	35	34	34	35				
	E	38	49	36	38	38	40				
GEID	APM	41	41	40	40	40	40				
	A										
	E										
FI-GEID	APM	41	41	33	33	33	33				
	A										
	E	39	49	35	32	34	37				
3SLS	APM	41	41	33	33	33	33				
	A	41	41	33	33	32	33				
	E	38	49	34	34	34	37				
FIML	A	42	42	34	33	33	34				
	E	40	49	35	33	34	37				
A2SLS	A	40	40	35	34	33	34	123			122
	E	38	53	35	37	39	40	135			166
AGEID	A										
	E	42	56	43	46	49	46	135			165
3SLSA-Rd	A	40	40	33	32	32	32	120			118
	E	39	52	34	35	34	37	131			155
FIMLA-Rd	A										
	E										
GEIDA	A										
	E	41	52	43	47	47	46	140	147	149	159
3SLSA	A	41	40	32	32	31	32	154	153	155	151
	E	41	53	35	36	34	37	149	146	164	170
FIMLA	A	43	42	34	33	33	34	161	160	163	158
	E	43	55	36	35	36	39	154	147	168	165

TABLE C4. Mean R2- and "d"-values of the CLID-specified structural form. Associated standard deviations are given within the parentheses.

	R2 relation 1		R2 relation 2		"d" relation 1		"d" relation 2	
OLS 2SLS	.945	(.014)	.943	(.018)	1.98	(.246)	1.97	(.287)
GEID FI-GEID 3SLS FIML	.936	(.018)	.934	(.022)	1.96	(.236)	1.98	(.288)
A2SLS	.939	(.017)	.938	(.022)	1.97	(.056)	1.98	(.052)
AGEID	.936	(.019)	.936	(.023)	1.97	(.056)	1.98	(.055)
3SLSA-Rd FIMLA-Rd	.938	(.017)	.936	(.022)	1.96	(.123)	1.96	(.108)
GEIDA	.937	(.018)	.936	(.022)	1.97	(.061)	1.98	(.056)
3SLSA FIMLA	.938	(.017)	.937	(.022)	1.97	(.062)	1.98	(.053)

TABLE D1. Mean square errors experiment D. $10^5(K^{-1}\underset{k}{\Sigma}(\hat{\theta}_{ijk} - \theta_i)^2)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1256	1702	322	271	350	252				
2SLS	407	196	322	156	217	145				
GEID	432	202	417	269	323	174				
FI-GEID	386	198	210	104	143	217				
3SLS	359	194	217	134	139	174				
FIML										
A2SLS	55	46	42	41	23	43	2852			1853
AGEID	61	45	64	58	33	60	4166			1977
3SLSA-Rd	50	43	33	35	19	39	2298			1561
FIMLA-Rd	56	45	35	34	21	37	4955			6456
GEIDA-Rd	110	109	84	96	109	104	5798			3506
3SLSA	59	46	39	46	19	44	6089	2183	3107	3875
FIMLA	67	49	40	47	22	44	10384	4071	4817	8791

TABLE D2. Mean biases experiment D. $10^4(K^{-1}\underset{k}{\Sigma}\hat{\theta}_{ijk} - \theta_i)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	898	1003	−150	−322	−319	−337				
2SLS	29	136	15	−72	−70	−58				
GEID	16	141	198	−133	−126	−36				
FI-GEID	11	62	62	17	21	−91				
3SLS	51	13	76	20	−12	−76				
FIML										
A2SLS	55	92	47	−39	−37	7	−843			−554
AGEID	24	66	70	−42	−45	23	−957			−576
3SLSA-Rd	98	72	36	−26	−16	6	−751			−609
FIMLA-Rd	28	41	33	−28	−18	10	−1571			−1749
GEIDA-Rd	61	17	58	−44	−29	38	−1347			−714
3SLSA	49	69	33	−34	−24	5	−1459	304	−247	−756
FIMLA	40	41	30	−38	−22	8	−2385	421	−350	−1706

TABLE D3. Square roots of the variances experiment D. The values are multiplied by 10^3.

A = according to the asymptotic formulae; the entires are means calculated over the 50 generated samples.

E = as obtained from the empirical sampling distributions.

		β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	A	39	43	39	38	41	42				
	E	68	84	55	41	50	37				
2SLS	A	52	53	44	43	45	47				
	E	64	43	55	39	47	38				
GEID	A	52	54	50	51	52	53				
	E	66	43	62	51	56	42				
FI-GEID	A	50	56	41	41	72	69				
	E	63	44	46	33	38	46				
3SLS	A	51	53	41	38	41	43				
	E	60	43	46	37	38	41				
FIML	A										
	E										
A2SLS	A	19	19	16	16	17	17	74			69
	E	23	19	20	20	15	21	148			126
AGEID	A										
	E	25	21	25	24	18	25	182			130
3SLSA-Rd	A	19	19	15	15	16	16	88			81
	E	22	20	18	19	14	20	133			109
FIMLA-Rd	A	19	20	16	16	16	17	113			109
	E	24	21	18	18	15	19	157			184
GEIDA-Rd	A										
	E	33	33	29	31	33	33	202			173
3SLSA	A	19	19	15	15	15	16	124	115	128	113
	E	24	21	20	21	14	21	201	146	176	184
FIMLA	A	19	20	16	16	16	16	153	151	160	148
	E	26	22	20	22	15	21	217	199	219	243

TABLE D4. Mean R2- and "d"-values of the CLID-specified structural form. Associated standard deviations are given within the parentheses.

OLS	.954	(.043)	.946	(.036)	.76	(.296)	.67	(.293)
2SLS	.942	(.056)	.935	(.051)	.53	(.303)	.46	(.267)
GEID	.941	(.057)	.935	(.051)	.60	(.375)	.48	(.269)
FI-GEID	.940	(.058)	.932	(.053)	.50	(.273)	.46	(.273)
3SLS	.941	(.055)	.934	(.052)	.51	(.277)	.45	(.275)
FIML								
A2SLS	.988	(.004)	.988	(.006)	1.81	(.326)	1.81	(.379)
AGEID	.988	(.004)	.988	(.006)	1.82	(.319)	1.81	(.389)
3SLSA-Rd	.981	(.005)	.987	(.006)	1.81	(.348)	1.79	(.359)
FIMLA-Rd								
GEIDA	.987	(.005)	.985	(.007)	1.87	(.299)	1.87	(.386)
3SLSA	.989	(.004)	.988	(.006)	1.83	(.314)	1.83	(.282)
FIMLA								

TABLE E1. Mean square errors experiment E. $10^5(K^{-1}\sum_k(\hat{\theta}_{ijk} - \theta_i)^2)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1206	1048	205	166	168	163				
2SLS	230	189	147	69	144	129				
GEID	231	191	188	84	183	185				
FI-GEID	227	198	124	105	138	93				
3SLS	228	189	126	86	134	98				
FIML	232	184	123	95	130	96				
A2SLS	213	140	120	87	144	106	1176			2619
AGEID	208	194	141	112	196	169	1231			2644
3SLSA-Rd	223	134	115	97	134	82	1092			1832
FIMLA-Rd										
GEIDA	238	185	156	110	202	154	2674	2449	4120	3243
3SLSA	251	163	105	106	161	84	2872	3330	4440	3659
FIMLA	230	191	106	107	157	93	3071	3196	4920	3844
FIMLA	209	188	105	94	133	87	1175		2730	

TABLE E2. Mean biases experiment E. $10^4(K^{-1}\sum_k\hat{\theta}_{ijk} - \theta_i)$

	β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	1200	963	−324	−330	−241	−249				
2SLS	106	−123	−65	−98	71	51				
GEID	106	−122	−79	−100	87	24				
FI-GEID	79	−155	−69	−81	38	91				
3SLS	105	−124	−65	−83	52	78				
FIML	97	−126	−76	−81	58	88				
A2SLS	182	63	−95	−118	21	8	230			782
AGEID	96	−37	−97	−123	65	5	250			768
3SLSA-Rd	133	22	−47	−63	28	59	−506			531
FIMLA-Rd										
GEIDA	89	−110	−73	−98	91	39	91	190	113	−274
3SLSA	96	−71	−58	−79	41	70	195	194	286	−327
FIMLA	56	−123	−64	−81	50	86	−24	162	338	−553
FIMLA	72	−128	−71	−83	56	81	−3		22	

TABLE E3. Square roots of the variances experiment E. The values are multiplied by 10^3.

A = according to the asymptotic formulae; the entries are means calculated over the 50 generated samples.

E = as obtained from the empirical sampling distributions.

		β_{12}	β_{21}	γ_{11}	γ_{12}	γ_{23}	γ_{24}	ρ_{11}	ρ_{12}	ρ_{21}	ρ_{22}
OLS	A	32	35	31	31	33	32				
	E	32	35	32	24	33	32				
2SLS	A	39	43	34	34	36	36				
	E	47	42	38	25	38	36				
GEID	A	39	43	39	39	41	40				
	E	47	42	43	27	42	43				
FI-GEID	A										
	E	48	42	35	32	37	29				
3SLS	A	39	43	32	32	34	34				
	E	47	42	35	28	37	31				
FIML	A										
	E										
A2SLS	A	37	41	32	32	35	34	121			123
	E	43	37	34	27	38	33	107			140
AGEID	A										
	E	45	44	37	31	44	42	109			145
3SLSA-Rd	A	37	41	31	31	33	33	119			121
	E	46	37	34	31	37	28	92			125
FIMLA-Rd	A										
	E										
GEIDA	A										
	E	48	42	39	32	44	39	165	172	205	180
3SLSA	A	37	41	30	30	33	32	150	145	161	150
	E	50	40	32	32	40	28	170	183	211	190
FIMLA	A	38	43	31	31	33	34	154	295	168	155
	E	48	42	32	32	40	30	174	180	217	190
FIMLA	A	38	43	31	31	33	34	134		145	
	E	46	42	32	30	36	29	108		165	

TABLE E4. Mean R2- and "d"-values of the CLID-specified structural form. Associated standard deviations are given within the parentheses.

	R2 relation 1		R2 relation 2		"d" relation 1		"d" relation 2	
OLS	.946	(.014)	.941	(.021)	1.66	(.173)	1.88	(.254)
2SLS								
GEID	.937	(.019)	.930	(.026)	1.56	(.185)	1.82	(.240)
FI-GEID	.937	(.019)	.929	(.027)	1.56	(.191)	1.80	(.244)
3SLS	.938	(.019)	.930	(.027)	1.56	(.188)	1.81	(.243)
FIML								
A2SLS	.943	(.016)	.935	(.024)	1.97	(.077)	1.96	(.053)
AGEID	.941	(.017)	.933	(.026)	1.97	(.077)	1.96	(.054)
3SLSA-Rd	.941	(.018)	.934	(.025)	1.83	(.136)	1.91	(.109)
FIMLA-Rd								
GEIDA	.942	(.017)	.936	(.024)	1.97	(.088)	1.96	(.076)
3SLSA	.943	(.017)	.937	(.023)	1.97	(.088)	1.96	(.078)
FIMLA								

TABLE I. Mean R2- and "d"-values of the reduced form experiment E. Associated standard deviations are given within the parentheses.

	R2 relation 1		R2 relation 2		"d" relation 1		"d" relation 2	
OLS	.777	(.060)	.774	(.078)	1.60	(.203)	1.69	(.235)
2SLS								
GEID	.803	(.046)	.802	(.062)	1.56	(.209)	1.67	(.229)
FI-GEID								
3SLS								
FIML								
A2SLS	.816	(.043)	.811	(.061)	1.86	(.101)	1.86	(.084)
AGEID								
3SLSA-Rd								
FIMLA-Rd								
GEIDA	.821	(.041)	.819	(.055)	1.96	(.092)	1.96	(.084)
3SLSA	.819	(.041)	.818	(.050)	1.96	(.093)	1.95	(.085)
FIMLA								

CHAPTER 8: ESTIMATION OF KLEIN'S MODEL I

8.1 The model

In chapter 7 the estimation methods developed were examined through a Monte Carlo experiment. The purpose of this chapter is to get an indication of the working of the methods for real world models. We will try to estimate Klein's model I. In the original version as given by Klein (1950) this model consists of three behavioral relations and three identities. The behavioral relations are a consumption equation, an investment equation and a wage equation. They are specified as follows:

$$C_t = a_{10} + a_{11}P_t + a_{12}W_t + a_{13}P_{t-1} + \delta_{1t} \tag{1a}$$

$$I_t = a_{20} + a_{21}P_t + a_{22}P_{t-1} + a_{23}K_{t-1} + \delta_{2t} \tag{1b}$$

$$W_t^0 = a_{30} + a_{31}F_t + a_{32}F_{t-1} + a_{33}(t-1931) + \delta_{3t} \tag{1c}$$

where the a's denote the parameters and C_t , I_t , .. $(t-1931)$ the variables. The identities are

$$Y_t + T_t = C_t + I_t + G_t \tag{2a}$$

$$Y_t = P_t + W_t \tag{2b}$$

$$K_t - K_{t-1} = I_t \tag{2c}$$

To estimate the system it is necessary to identify the endogenous variables. Three-C_t aggregate consumption, I_t net investment and W_t^0 private wage bill—are given by Klein as the left hand variables of the behavioral relations. Three-Y_t national income, P_t profits and K_t end of year capital stock—are obtained from the identities, since they are functions of endogenous variables. Total wage bill W_t is a sum of private wage bill W_t^0, which is endogenous, and government wage bill W_t^{00}. This means that W_t is endogenous. The variable can be defined through an identity. Likewise for F_t which is private product, or in other words national income Y_t with indirect taxes T_t added and government wage bill W_t^{00} subtracted. We now have obtained the eight endogenous variables of the system.

The behavioral relations above were written with a dependent variable on the left hand side. We find it convenient to write the whole system, behavioral relations and identities, asymmetrically with each endogenous variable appearing exactly one time on the left hand side, cf. section 2.1. This is obtained by specifying the identities e.g. as

$$Y_t = C_t + I_t + G_t - T_t \tag{3a}$$

$$P_t = Y_t - W_t^0 - W_t^{00} \tag{3b}$$

$$K_t = K_{t-1} + I_t \tag{3c}$$

$$W_t = W_t^0 + W_t^{00} \tag{3d}$$

$$F_t = Y_t + T_t - W_t^{00} \tag{3e}$$

The equations (1a–c) and (3a–e) are our definition of the basic system. The variables not yet defined are G_t government expenditure and t time trend. The subscript t–1 is attached to lagged variables.

8.2 Results

The methods of table 7.2.1 were applied to the system defined above using the 21 observations 1921–1941 given by Klein. The parameter estimates with the associated standard deviations, as calculated from the asymptotic formulae, are given in table 1. The residual characteristics are shown in tables 2–4. Their definitions were given in subsection 7.3.1. As regards the group II and III methods the characteristics were calculated for the residuals of the transformed system, using the 20 observations then available.

The "d"-values of the structural form for the group I methods OLS-FIML, indicate that the more sophisticated methods should perhaps be applied with only one non-zero element ρ_{11}–or no non-zero element–in R. The "d"-values of the first relation are contained in the inconclusive region of the Durbin-Watson test, while the other are contained in the region of non-rejection. But the purpose is to examine the applicability of the estimation methods and therefore the main alternatives are a diagonal and a full auto-regressive structure matrix R.

As is seen in the tables, two sets of FIMLA estimates were obtained. Both correspond to local maxima and are earlier given by Hendry (1971, 1974). The likelihood of the second is the highest; the function value L is .855[1]. This solution was attained only when the starting values of the iterative procedure were chosen in its very neighbourhood. The first solution–denoted FIMLA_I–was obtained starting either with all parameters zero, or the FIML-estimates or the FIMLA-Rd estimates; the function value L is .719.

[1] $L = -.5 \log\det \Omega + \log\det \beta_I$.

A systematic search for several extrema was not performed. Generally we were satisfied with finding one solution. For some estimators even this limited aim was hard to attain. The iterative methods used to calculate AGEID and 3SLSA estimates turned out to be very sensitive to the choice of starting values. AGEID and 3SLSA belong to the GLS-approach methods defined in section 5.1. GEIDA, the "calculated residuals as regressors" approach method, worked—as expected—better in this respect. The AGEID estimates had to be calculated in three steps. The starting point for an initial AGEID estimation, with ρ_{22} and ρ_{33} *a priori* equal to zero, was the Y^{**}'s obtained from the restricted reduced form derived from the corresponding set of A2SLS estimates. The obtained solution was the starting point in a second estimation round with ρ_{33} *a priori* equal to zero. The final AGEID estimates are shown in the tables. They were found with the estimates of the second round as starting values. The number of iterations and the CPU-time given in table 5 refer to this third round only. With other starting procedures the method did not converge.

It was still harder to find 3SLSA estimates. For the case with R full some different starting values—i.e. different initial disturbance covariance matrices—were tried, but within the iteration limit of 200 iterations no convergence was reached. Not even with a fixed "known" covariance matrix did the method work; e.g. was the disturbance covariance matrix of the $FIMLA_I$ solution tried, but without success. The case with a diagonal R matrix worked better. With a known "fixed" covariance matrix estimates were obtained rather easily. When the 3SLSA estimator was extended by adding equations for the estimation of the covariance matrix, according to comment 5.4.2, no final estimates were reached unless ρ_{33} was or ρ_{22} and ρ_{33} were *a priori* zero. The 3SLSA-Rd results for the case with ρ_{33} restricted are given in the tables.

Let us discuss the estimates somewhat. The signs of the coefficients in table 1 are not always as expected. The coefficient for P_t in the consumption function should according to economic theory have positive sign. FI-GEID, FIML, AGEID and GEIDA give negative values, but the associated standard deviations are on the other hand rather large. The behavior for the investment function is poorer. We expect positive signs for P_t here too, but FI-GEID and to some extent FIML, 3SLSA-Rd and $FIMLA_I$ yield large negative values. The R2-value in table 2 is for the second relation as estimated by FI-GEID –.161. The corresponding value for the reformulated structural form is .747.

A general impression is that the methods accounting for serial correlation

do not give estimates more unrealistic than the standard methods. The increase in fit, going from the methods of group I to those of group III, is demonstrated in table 3. When interpreting the figures it should be kept in mind that the R2-values are not corrected for degrees of freedom.

TABLE 1. Klein's model I. Estimated structural parameters and associated standard deviations (within parentheses).
a) THE CONSUMPTION FUNCTION

Variable	P_t	W_t	P_{t-1}	δ_{1t-1}	δ_{2t-1}	δ_{3t-1}	const.
OLS	.193 (.082)	.796 (.036)	.090 (.082)				16.24
2SLS	.017 (.118)	.810 (.040)	.216 (.107)				16.55
GEID	.022 (.111)	.794 (.041)	.256 (.109)				16.50
FI-GEID	−.466 (.565)	.817 (.060)	.542 (.428)				19.09
3SLS	.125 (.108)	.790 (.038)	.163 (.100)				16.44
FIML	−.232 (.312)	.802 (.036)	.386 (.217)				18.34
GEIDA-Rd	.016	.705	.154	.694			22.39
A2SLS	.040 (.159)	.770 (.068)	.157 (.119)	.419 (.028)			18.88
AGEID	−.028	.804	.265	.210			16.81
3SLSA-Rd[2]	.128 (.095)	.789 (.050)	.155 (.083)	.290 (.155)			16.55
3SLSA-Rd[3]	.115 (.108)	.796 (.058)	.129 (.100)	.357 (.240)			16.91
FIMLA-Rd[5]	.204 (.088)	.632 (.124)	.114 (.064)	.73 (.23)			23.05
GEIDA	−.054	.765	.262	.536	−.212	−.143	18.96
$FIMLA_{I_5}^{4/5}$	.089 (.169)	.730 (.091)	.224 (.159)	.72 (.38)	−.23 (.20)	.33 (.49)	18.70
$FIMLA_{II}$	.152 (.082)	.547 (.118)	.014 (.072)	.90 (.14)	.11 (.18)	−.22 (.22)	29.56

b) THE INVESTMENT FUNCTION

Variable	P_t	P_{t-1}	K_{t-1}	δ_{1t-1}	δ_{2t-1}	δ_{3t-1}	const.
OLS	.480 (.087)	.333 (.091)	−.112 (.024)				10.13
2SLS	.150 (.173)	.616 (.163)	−.158 (.036)				20.28
GEID	.089 (.167)	.665 (.165)	−.166 (.038)				22.19
FI-GEID	−.991 (.209)	1.386 (.177)	−.185 (.027)				32.40
3SLS	.013 (.162)	.756 (.153)	−.195 (.033)				28.18
FIML	−.801 (.491)	1.052 (.352)	−.148 (.030)				27.26

GEIDA-Rd	.055	.686	−.189		−.058		27.18
A2SLS	−.001 (.250)	.730 (.221)	−.198 (.055)		−.042 (.218)		29.07
AGEID	.032	.698	−.194		−.034		28.36
3SLSA-Rd[2]	−.314 (.311)	.969 (.273)	−.255 (.068)		−.584 (.192)		41.97
3SLSA-Rd[3]	−.082 (.242)	.794 (s.213)	−.217 (.054)		−.041 (.197)		33.24
FIMLA-Rd[5]	.104 (.217)	.634 (.185)	−.176 (.051)		−.03 (.22)		24.49
GEIDA	−.108	.794	−.158	.531	−.267	−1.021	21.60
$FIMLA_{I_5}^{4/5}$	−.378 (.433)	1.043 (.369)	−.240 (.074)	.27 (.91)	.10 (.35)	−1.64 (.99)	38.63
$FIMLA_{II}$	.297 (.100)	.583 (.118)	−.016 (.023)	−.40 (.21)	−.39 (.20)	−1.26 (.31)	−11.14

c) THE WAGE EQUATION

Variable	F_t	F_{t-1}	t−1931	δ_{1t-1}	δ_{2t-1}	δ_{3t-1}	const.
OLS	.440 (.029)	.146 (.034)	.130 (.029)				1.50
2SLS	.439 (.036)	.147 (.039)	.130 (.029)				1.50
GEID	.401 (.046)	.171 (.051)	.162 (.038)				2.39
FI-GEID	.331 (.099)	.241 (.118)	.156 (.083)				2.55
3SLS	.401 (.032)	.181 (.034)	.150 (.028)				1.78
FIML	.234 (.049)	.285 (.045)	.235 (.034)				5.79
GEIDA-Rd	.432	.136	.125			−.581	2.58
A2SLS	.433 (.030)	.145 (.034)	.117 (.025)			−.134 (.027)	2.02
AGEID	.436	.129	.122			−.098	2.80
3SLSA-Rd[2]	.393 (.027)	.171 (.027)	.132 (.026)				2.96
3SLSA-Rd[3]	.412 (.029)	.157 (.028)	.124 (.026)			.036 (.026)	2.55
FIMLA-Rd[5]	.456 (.056)	.123 (.054)	.134 (.031)			.21 (.21)	1.93
GEIDA	.460	.065	.078	.435	−.067	.351	5.23
$FIMLA_{I_5}^{4/5}$	.395 (.030)	.169 (.030)	.134 (.031)	−.01 (.42)	.12 (.15)	−.38 (.44)	2.90[6]
$FIMLA_{II}$	.548 (.040)	−.022 (.031)	.075 (.052)	.23 (.14)	.13 (.13)	.49 (.17)	5.33[7]

TABLE 2. Klein's model I. Residual characteristics for the CLID-specified structural form: R2 the multiple correlation coefficient squared and "d" the Durbin-Watson statistic.

	R2			"d"		
	rel 1	rel 2	rel 3	rel 1	rel 2	rel 3
OLS	.981	.931	.987	1.37	1.81	1.96
2SLS	.977	.885	.987	1.49	2.09	1.96
GEID	.976	.866	.986	1.51	2.07	2.09
FI-GEID	.916	−.161	.979	1.34	1.52	2.07
3SLS	.980	.826	.986	1.42	2.00	2.16
FIML	.953	−.063	.952	1.37	1.24	1.49
GEIDA-Rd	.973	.868	.985	1.92	2.02	1.59
A2SLS	.975	.850	.989	1.78	2.03	2.07
AGEID	.970	.861	.988	1.75	2.03	2.05
3SLSA-Rd[2]	.978	.706	.986	1.70	1.86	2.29
3SLSA-Rd[3]	.978	.820	.987	1.75	1.99	2.33
FIMLA-Rd	.980	.881	.987	1.91	2.06	2.18
GEIDA	.971	.880	.989	1.89	2.28	1.95
$FIMLA_I^4$	.978	.785	.989	1.90	2.18	2.23
$FIMLA_{II}$	.972	.953	.985	1.84	2.58	1.78

TABLE 3. Klein's model I. Residual characteristics for the reduced form: R2, the multiple correlation coefficient squared.

	rel 1	rel 2	rel 3	rel 4	rel 5	rel 6	rel 7	rel 8
OLS	.825	.632	.887	.790	.497	.945	.921	.785
2SLS	.913	.833	.928	.902	.786	.975	.950	.900
GEID	.918	.835	.930	.905	.788	.976	.951	.903
FI-GEID	.883	.747	.890	.858	.736	.963	.924	.854
3SLS	.914	.831	.929	.903	.785	.975	.950	.901
FIML	.884	.684	.868	.842	.724	.953	.908	.839
GEIDA-Rd	.942	.851	.959	.930	.806	.972	.971	.932
A2SLS	.935	.850	.949	.920	.815	.972	.964	.923
AGEID	.928	.850	.944	.914	.806	.971	.960	.917
3SLSA-Rd[2]	.930	.829	.934	.912	.814	.967	.954	.914
3SLSA-Rd[3]	.929	.846	.941	.915	.815	.971	.958	.918
FIMLA-Rd	.929	.853	.947	.917	.804	.972	.963	.920
GEIDA	.946	.905	.963	.942	.854	.982	.974	.944
$FIMLA_I^4$	.946	.894	.957	.939	.856	.980	.970	.940
$FIMLA_{II}$	.907	.872	.953	.902	.762	.976	.967	.905

TABLE 4. Klein's model I. Residual characteristics for the reduced form: "d", the Durbin-Watson statistic.

	rel 1	rel 2	rel 3	rel 4	rel 5	rel 6	rel 7	rel 8
OLS	1.23	1.57	1.36	1.38	1.47	1.57	1.36	1.33
2SLS	1.62	2.02	1.99	1.81	1.68	2.02	1.99	1.81
GEID	1.72	2.04	2.09	1.88	1.72	2.04	2.09	1.88
FI-GEID	1.43	1.66	1.75	1.53	1.42	1.66	1.75	1.53
3SLS	1.59	2.00	2.05	1.79	1.63	2.00	2.05	1.79
FIML	1.22	1.27	1.35	1.23	1.25	1.27	1.35	1.23
GEIDA-Rd	2.17	2.01	2.25	2.14	1.88	2.01	2.25	2.14
A2SLS	1.98	2.03	2.36	2.04	1.82	2.03	2.36	2.04
AGEID	1.98	2.03	2.31	2.04	1.84	2.03	2.41	2.04
3SLSA-Rd[2]	1.90	1.89	2.22	1.94	1.77	1.89	2.22	1.94
3SLSA-Rd[3]	1.91	2.00	2.33	2.00	1.80	2.00	2.33	2.00
FIMLA-Rd	2.05	2.03	2.40	2.06	1.87	2.03	2.40	2.06
GEIDA	2.37	2.29	2.67	2.38	2.66	2.29	2.67	2.38
$FIMLA_I$[4]	2.18	2.22	2.46	2.24	2.03	2.22	2.46	2.24
$FIMLA_{II}$	1.98	2.27	2.39	2.07	1.90	2.27	2.39	2.06

TABLE 5. Klein's model I. The behavior of the numerical methods; CPU-time and number of iterations required. The convergence criteria are given in subsection 7.3.4.

	CPU-time in seconds	Number of iterations
GEID	7	17
FI-GEID	43	190
FIML		
GEIDA-Rd	6	28
A2SLS		17
AGEID	16	38
3SLSA-Rd[2]	22	50
3SLSA-Rd[3]	14	28
FIMLA-Rd	19	
GEIDA	13	59
$FIMLA_I$[4]	21	
$FIMLA_{II}$		

[2] ρ_{33} is a priori zero.

[3] The weighting matrix Ω is estimated in the first iteration.

[4] $FIMLA_{II}$ is the solution with the largest likelihood.

[5] The standard deviations are obtained utilizing the E-matrix approximating the inverse of the matrix of the second order derivatives.

[6] Hendry (1971) obtained 1.42.

[7] Hendry (1971) obtained 4.50.

REFERENCES

Ågren, A. (1972): *Extensions of the Fix-Point Method. Theory and Applications.* FD-thesis. Selected publications 26. University Institute of Statistics, Uppsala.

Ågren, A. (1973): A Short Note on the GEID Specification and the GEID Estimator. Chapter 3 of Lyttkens & Wold (1973).

Ågren, A. (1974): On the Specification and Estimation of GEID-Systems. *Research report* 74–9. *University Institute of Statistics, Uppsala.*

Ågren, A. and H.O. Wold (1969): On the Structure and Estimation of General Interdependent Systems. In *Multivariate Analysis II,* edited by P.R. Krishnaiah. Academic Press, New York.

Amemiya, T. (1966): Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Parameters. *Econometrica 34*, pp. 283–306.

Anderson, T.W. and H. Rubin (1949): Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics* 20, pp. 46–63.

Anderson, T.W. and H. Rubin (1950): The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Compete System of Stochastic Equations. *Annals of Mathematical Statistics* 21, pp. 570–582.

Ball, R.J., editor (1973): *The International Linkage of National Economic Models.* North-Holland, Amsterdam.

Basmann, R.L. (1957): A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica* 25, pp. 77–83.

Bergström, R. (1974): *Studies in the Estimation of Interdependent Systems especially the Fix-Point and Iterative Instrumental Variables Methods.* FD-thesis. Selected publications 33. University Institute of Statistics, Uppsala.

Blattberg, R.C. (1973): Evaluation of the Power of the Durbin-Watson Statistic for Non-First Order Serial Correlation Alternatives. *Review of Economics and Statistics* LV, pp. 508–517.

Bodin, L. (1974): *Recursive Fix-Point Estimation Theory and Applications.* FD-thesis. Selected Publications 32. University Institute of Statistics, Uppsala.

Brundy, J. and D. Jorgenson (1971): Efficient Estimation of Simultaneous Equations by Instrumental Variables. *Review of Economics and Statistics* LIII, pp. 207–224.

Chow, G.C. and R.C. Fair (1973): Maximum Likelihood Estimation of Linear Equation Systems with Autoregressive Residuals. *Annals of Economic and Social Measurement* 2, pp. 17–28.

Christ, C.F. (1966): *Econometric Models and Methods.* Wiley, New York.
Cochrane, D. and G.H. Orcutt (1950): Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. *Journal of the American Statistical Association* 44, pp. 32–61.
Cramér, H. (1946): *Mathematical Methods of Statistics.* Princeton University Press, Princeton.
Dhrymes, Ph.J. (1970): *Econometrics, Statistical Foundations and Applications.* Harper & Row, Evanston.
Dhrymes, Ph.J. (1971): *Distributed Lags.* Holden-Day, San Francisco.
Dhrymes, Ph.J., R. Berner and D. Cummins (1974): A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equation Models with Autocorrelated Errors. *Econometrica* 42, pp. 313–332.
Dhrymes, Ph.J. and H. Erlat (1974): Asymptotic Properties of Full Information Estimators in Dynamic Autoregressive Simultaneous Equation Models. *Journal of Econometrics* 2, pp. 247–259.
Diestler, M. (1975): z-Transform and Identification of Linear Econometric Models with Autocorrelated Errors. *Metrika* 22, pp. 13–25.
Durbin, J. (1957): Testing for Serial Correlation in Systems of Simultaneous Regression Equations. *Biometrika* 44, pp. 370–377.
Durbin, J. (1969): Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least-Squares Residuals. *Biometrika* 56, pp. 1–16.
Durbin, J. (1970A): Testing for Serial Correlation in Least-Squares Regression when some of the Regressors are Lagged Dependent Variables. *Econometrica* 38, pp. 410–421.
Durbin, J. (1970B): An Alternative to the Bounds Test for Testing for Serial Correlation in Least Squares Regression. *Econometrica* 38, pp. 422–429.
Durbin, J. and G.S. Watson (1950): Testing for Serial Correlation in Least Squares Regression I. *Biometrika* 37, pp. 409–428.
Durbin, J. and G.S. Watson (1951): Testing for Serial Correlation in Least Squares Regression II. *Biometrika* 38, pp. 159–178.
Durbin, J. and G.S. Watson (1971): Testing for Serial Correlation in Least Squares Regression III. *Biometrika* 58, pp. 1–20.
Edgerton, D. (1973): *Nonlinear Interdependent Systems.* FD-thesis. Selected Publications 30. University Institute of Statistics, Uppsala.
Evans, M.K. and L.R. Klein (1968): *The Wharton Econometric Forecasting Model.* Economics Research Unit, University of Philadelphia, Philadelphia.
Fair, R.C. (1970): The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors. *Econometrica* 38, pp. 507–516.

Fair, R.C. (1972A): *A Short-Run Forecasting Model of the United States' Economy*. D.C. Heath & Co., Lexington.

Fair, R.C. (1972B): Efficient Estimation of Simultaneous Equations with Autoregressive Errors by Instrumental Variables. *Review of Economics and Statistics* LIV, pp. 444–449.

Fair, R.C. (1973): A Comparison of Alternative Estimators of Macroeconomic Models. *International Economic Review* 14, pp. 261–277.

Fair, R.C. (1974): An Evaluation of a Short-Run Forecasting Model. *International Economic Review* 15, pp. 285–303.

Fisher, F. (1966): *The Identification Problem in Econometrics.* McGraw-Hill, New York.

Fisk, P.R. (1967): *Stochastically Dependent Equations.* Griffin Statistical Monograph 21, London.

Fletcher, R. and J. Powell (1963): A Rapidly Convergent Descent Method for Minimization. *The Computer Journal* 6, pp. 163–168.

Goldberger, A.S. (1965): *Econometric Theory.* Wiley, New York.

Goldfeld, S.M. and R.E. Quandt (1972): *Nonlinear Methods in Econometrics.* North-Holland, Amsterdam.

Grether, D.M. and G.S. Maddala (1972): On the Asymptotic Properties of Some Two Step Procedures for Estimating Distributed Lag Models. *International Economic Review* 13, pp. 737–744.

Gruvaeus, G. and K.G. Jöreskog (1970): A Computer Program for Minimizing a Function of Several Variables. *Research Bulletin* 70–14, *Educational Testing Service, Princeton.*

Guilkey, D.K. and P. Schmidt (1973): Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors. *Journal of the American Statistical Association* 68, pp. 642–647.

Haavelmo, T. (1943): The Statistical Implications of a System of Simultaneous Equations. *Econometrica* 11, pp. 1–12.

Habibagahi, H. and J.L. Pratschke (1972): A Comparison of the Power of the von Neumann ratio, Durbin-Watson and Geary Tests. *Review of Economics and Statistics* LIV, pp. 179–186.

Hart, B.I. (1942): Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance and Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance. *Annals of Mathematical Statistics* 13, pp. 207–214 and 445–447.

Hendry, D.F. (1971): Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process. *International Economic Review* 12, pp. 257–272.

Hendry, D.F. (1974): Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process: A Correction. *International Economic Review* 15, p. 260.

Hickman, B.G., editor (1972): *Econometric Models of Cyclical Behavior.* Studies in Income and Wealth, 36. National Bureau of Economic Research. Columbia University Press, New York.

IMSL Library 1 Reference Manual (1973): International Mathematical and Statistical Libraries, Houston.

Johnston, J. (1972): *Econometric Methods.* McGraw-Hill, New York.

Jöreskog, K.G. (1973): A General Method for Estimating a Linear Structural Equation System. In *Structural Equation Models in the Social Sciences,* edited by A.S. Goldberger and O.D. Duncan. Seminar Press, New York.

Klein, L.R. (1950): *Economic Fluctuations in the United States* 1921–1941. Cowles Commission Monograph 11. Wiley, New York.

Klein, L.R. (1969): Estimation of Interdependent Systems in Macro-Econometrics. *Econometrica* 37, pp. 171–192.

Klein, L.R. and A.S. Goldberger (1955): *An Econometric Model of the United States* 1929–1952. North-Holland, Amsterdam.

Koerts, J. and A.P.J. Abrahamse (1968): On the Power of the BLUS Procedure. *Journal of the American Statistical Association* 63, pp. 1227–1236.

Koopmans, T.C., editor (1950): *Statistical Inference in Dynamic Economic Models.* Cowles Commission Monograph 10. Wiley, New York.

Koopmans, T.C., Rubin, H. and R.B. Leipnik (1950): Measuring the Equation System of Dynamic Economics. Chapter II of Koopmans (1950).

Liu, Ta-Chung (1969): A Monthly Recursive Econometric Model of the United States: A Test of Feasibility. *Review of Economics and Statistics* LI, pp. 1–13.

Lyttkens, E. (1964): Standard Errors of Regression Coefficients by Autocorrelated Residuals. In *Econometric Model Building,* edited by H.O. Wold. North-Holland, Amsterdam.

Lyttkens, E. (1967): On the Maximum Likelihood Estimation of an Asymmetrically Specified GEID System. *Seminar paper. University Institute of Statistics, Uppsala.*

Lyttkens, E. (1970): Sections of Mosbaek and Wold (1970).

Lyttkens, E. (1973): The Fix-Point Method for Estimating Interdependent Systems with the Underlying Model Specification. *Journal of the Royal Statistical Society, Series A* 136, pp. 353–394.

Lyttkens, E. and H.O. Wold, editors (1973): FP, IIV and related approaches to ID systems. Group report presented at *The European Meeting of the Econometric Society, Oslo.* Also *Research Report* 73–8 *University Institute of Statistics, Uppsala* and *Report No. 6 Statistics Department, University of Gothenburg.*

Madansky, A. (1964): On the Efficiency of Three Stage Least Squares Estimation. *Econometrica* 32, pp. 51–56.

Maddala, G.S. (1971): Generalized Least Squares with an Estimated Variance Covariance Matrix, *Econometrica* 39, pp. 23–34.

Malinvaud, E. (1961): Estimation et Prévision dans les Modèles Économiques Autorégressifs. *Revue de l'Institute International de Statistique* 29, pp. 1–32.

Malinvaud, E. (1970): *Statistical Methods of Econometrics.* North-Holland, Amsterdam.

Mallella, P. and A. Rao (1971): Identifiability with Lagged Dependent Variables and Auto-Correlated Disturbances. Abstract of paper presented at Second World Congress of Econometric Society. *Econometrica* 39, p. 201.

Mosbaek, E.J. and H.O. Wold (1970): *Interdependent Systems, Structure and Estimation.* With contributions by E. Lyttkens, A. Ågren and L. Bodin. North-Holland, Amsterdam.

Nerlove, M. and K. Wallis (1966): Use of the Durbin-Watson Statistic in Inappropriate Situations. *Econometrica* 34, pp. 235–238.

Neudecker, H. (1969): Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. *Journal of the American Statistical Association* 64, pp. 953–963.

Neumann, J. von (1941): Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *Annals of Mathematical Statistics* 12, pp. 367–395.

Norlén, U. (1965): Studies in Nonlinear Estimation of Multirelational Models. *Mimeographed FL-thesis. University Institute of Statistics, Uppsala.*

Parks, R. (1967): Efficient Estimation of a System of Regression Equations when Disturbances are both Serially and Contemporaneously Correlated. *Journal of the American Statistical Association* 62, pp. 500–509.

Phillips, A.W. (1966): Estimation of Systems of Difference Equations with Moving Average Disturbances. Paper read at *The San Francisco meeting of the Econometric Society,* December 1966.

Powell, J. (1971): Recent Advances in Unconstrained Optimization. *Mathematical Programming* 1, pp. 26–57.

Rao, P. and Z. Griliches (1969): Small Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors. *Journal of the American Statistical Association* 64, pp. 253–272.

Sargan, J.D. (1961): The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals. *Econometrica* 29, pp. 414–426.

Sargan, J.D. (1964): Wages and Prices in the United Kingdom. In *Econometric Analysis for National Economic Planning; Colston Papers* XVI. Butterworths, London.

Schönfeld, P. (1971A): *Methoden der Ökonometrie II.* Franz Vahlen, Berlin.

Schönfeld, P. (1971B): A Useful Central Limit Theorem for m-Dependent Variables. *Metrika* 17, pp. 116–128.

Selén, J. (1973A): Autocorrelation in Interdependent Econometric Models; Specification of REID and GEID Systems. *Seminar Paper, University Institute of Statistics, Uppsala.*

Selén, J. (1973B): Autocorrelated Residuals in ID-Systems. Chapter 8 of Lyttkens & Wold (1973).

Selén, J. (1974): On the Properties of Some Estimation Methods for Regression Models Containing Autocorrelated Error Terms. *Research Report 74–6. University Institute of Statistics, Uppsala.*

Siegel, S. (1956): *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, New York.

Smith, K.R. (1973): *Monte Carlo Methods. Their Role for Econometrics.* D.C. Heath & Co, Lexington.

Sowey, E.R. (1973): A Classified Bibliography of Monte Carlo Studies in Econometrics. *Journal of Econometrics* 1, pp. 377–397.

Summers, R.M. (1965): A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators. *Econometrica* 33, pp. 1–41.

Theil, H. (1958): *Economic Forecasts and Policy.* North-Holland, Amsterdam.

Theil, H. (1965): The Analysis of Disturbances in Regression Analysis. *Journal of the American Statistical Association* 60, pp. 1067–1079.

Theil, H. (1968): A Simplification of the BLUS Procedure for Analyzing Regression Disturbances. *Journal of the American Statistical Association* 63, pp. 242–281.

Theil, H. (1971): *Principles of Econometrics.* North-Holland, Amsterdam.

Theil, H. and A.L. Nagar (1961): Testing the Independence of Regression Disturbances. *Journal of the American Statistical Association* 56, pp. 793–806.

Watson, G.S. (1955): Serial Correlation in Regression Analysis I. *Biometrika* 42, pp. 327–341.

Watson, G.S. and E.J. Hannan (1956): Serial Correlation in Regression Analysis II. *Biometrika* 43, pp. 436–445.

Wickens, M.R. (1969): The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equation Systems with Autocorrelated Errors. *Econometrica* 37, pp. 651–659.

Wilks, S.S. (1962): *Mathematical Statistics.* Wiley, New York.

Wold, H.O. (1965): A Fix-Point Theorem with Econometric Background I–II. *Arkiv för Matematik* 6, pp. 209–239.

Wold, H.O. (1966): Nonlinear estimation by iterative least squares procedures. In *Research Papers in Statistics. Festschrift for J. Neyman,* edited by F.N. David. Wiley, New York.

Wold, H.O. and E. Lyttkens, editors (1969): Nonlinear Iterative Partial Least Squares (NIPALS) Estimation Procedures. *Bulletin de l'Institute International de Statistique* 53, pp. 29–51.

Zellner, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57, pp. 300–313.

Zellner, A. and H. Theil (1962): Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica* 30, pp. 54–78.

LiberTryck Stockholm 1975